



Universidad de Granada
Facultad de Comunicación y
Documentación



Universidad de La Habana
Facultad de Comunicación

Título

**UH-WEB: PROPUESTA DE DISEÑO DE UN CMS SEMÁNTICO
PARA LA UNIVERSIDAD DE LA HABANA**

Tesis Doctoral

Autora

MsC. Yorbelis Rosell León

Directores

Dr. José Antonio Senso Ruiz

Dr. Amed Abel Leiva Mederos

Granada, Septiembre 2016

Editor: Universidad de Granada. Tesis Doctorales

Autora: Yorbelis Rosell León

ISBN: 978-84-9163-044-9

URI: <http://hdl.handle.net/10481/44304>

Exergo

... “el bibliotecólogo debe también ser un agente de cambio social y motor del desarrollo”...¹

¹Sandí, M. [et al.] (2011). Modelo de relaciones transdisciplinarias para el diseño curricular en Ciencias Bibliotecológicas y de la Información. e-Ciencias de la Información. 1(1)

Dedicatoria

*A mi pequeña Isabella, inundaste de luz mi vida con sólo mirarme a
los ojos, minutos después de nacer*

A Hilda y Agustín, los mejores padres que un hijo pueda tener.

Los quiero

Agradecimientos

Quisiera agradecer a todas las personas que dejaron una huella a lo largo de las diferentes etapas de mi formación. Soy consciente que he llegado hasta aquí gracias a ellos. De manera muy especial mi más sincera gratitud:

A mis tutores José Senso y Amed Leiva, quienes más que directores han sido amigos por apostar por mí en este proyecto, la ayuda oportuna, la paciencia y la confianza.

Al equipo del grupo de investigación sobre la web semántica de la UCI, en especial a Yusniel Hidalgo y a Leandro Tabares, pues compartir conocimiento fue de invaluable ayuda en el desarrollo de esta investigación.

A Gloria Ponjuan, por extender su mano cuando más lo necesitaba, sus consejos y el cariño.

A mi familia por estar siempre, hacerme reír en los días más difíciles y sostener mi mano en momentos de flaqueza.

A Rafa Olivares (Bibliotecario de Granada) por su trabajo desinteresado y toda la bibliografía que me hizo llegar.

A mis compañeros del Departamento de Ciencias de la Información en FCOM, el Rectorado y el DICT por estar presentes y seguir de cerca mi crecimiento profesional, en especial al Rector Dr. Gustavo Cobreiro y a la Directora Dr. Yohannis Martí por el apoyo para el feliz término de esta investigación.

A mis amigas Grisel del Toro, Leydis Fuentes y Yanelis González por sus charlas, ayuda y buenos deseos desde la distancia, pero siempre cerca.

A todos los profesores del Doctorado por los conocimientos transmitidos.

A todos los que vivimos juntos esta etapa vital...

Muchas gracias

Resumen

La Universidad de La Habana (UH), se ha proyectado para que sus productos web y la información contenida y gestionada en ellos se integren de manera eficiente. Los CMS actualmente utilizados (esencialmente Drupal y Joomla), aunque permiten crear y mantener los sitios web con facilidad, no admiten el manejo integral de datos heterogéneos como los que se demandan en la UH, lo que incide de directamente en la interoperabilidad entre sus sistemas y, por consecuencia, en la gestión de la información. Se hace necesario, asumir un sistema que permita describir ontologías específicas que combinen gran variedad de propiedades en los datos, inferencias, así como el enlazado y manejo de datos heterogéneos y datos semánticos.

La propuesta de un CMS semántico para la institución aportaría cambios positivos a la coherencia entre los procesos de gestión de información y las políticas trazadas para aportar descripciones explícitas de los recursos en la Web para facilitar la interpretación de los documentos y de realizar procesos inteligentes de captura y tratamiento de información. Abriría la posibilidad de construcción de sitios interoperables entre sí, logrando unificación y reutilización de la información.

En la presente investigación se exponen las pautas metodológicas que sustentan el diseño de un Sistema Gestor de Contenido Semántico (S-CMS) aplicado a la UH. Como parte de la propuesta se realiza una introducción teórica al tema, así como la presentación del marco metodológico y marco lógico que sustentan la propuesta. Del modelo de CMS UH-WEB se presentan: principios y requerimientos, objetivos, requisitos de diseño, arquitectura y la ontología del CMS ideado.

Tabla de Contenido

INTRODUCCIÓN.....	25
Introducción	26
Escenario de investigación	28
Problema de investigación	30
Pregunta de investigación.....	30
Sistema de objetivos.....	30
Antecedentes	31
Alcance y novedad de los resultados	31
Limitaciones del estudio	32
Estructura Capitular	33
Referencias bibliográficas	35
CAPÍTULO I	36
CAPÍTULO 1: Marco teórico referencial: Los Sistemas Gestores de Contenido (CMS)	37
1.1.Sistemas de Gestión de Contenidos: análisis terminológico	37
1.2.Origen e historia de los CMS.....	42
1.3.Estructura de los CMS: modelación conceptual y principales componentes	50
1.4.Tipologías de CMS.....	60
1.5.¿Por qué estudiar los CMS desde las CI?	67
1.6.Referencias bibliográficas.....	73
CAPÍTULO II.....	76
2.1.Web semántica: Análisis terminológico	77
2.2.Origen e historia de la Web semántica	82
2.3.Estructura de la Web semántica	88
2.4.Anotación semántica	93

2.5.Estándares y lenguajes de marcado	95
2.5.1. XML: Extensible Markup Language	97
2.5.2. RDF: Resource Description Framework	100
2.5.3. OWL: Web Ontology Language	108
2.6.Ontologías y Web semántica.....	115
2.6.1. Ontología: conceptualización del término	117
2.6.2. Tipos de ontologías	121
2.6.3. Metodologías para la construcción de ontologías	123
2.6.4. Herramientas para la construcción de ontologías	130
2.6.5. Lenguajes para la construcción de ontologías	131
2.6.6. Principales ontologías reutilizables	135
2.6.6.1. SKOS(http://www.w3.org/2004/skos/core)	137
2.6.6.2. BIBO (http://bibliontology.com/).....	146
2.6.6.3. VIVO (http://vivoweb.org/ontology/core)	149
2.6.6.4. FOAF (http://www.foaf-project.org/).....	153
2.6.6.5. Geopolitical Ontology (http://aims.fao.org/aos/geopolitical.owl)	156
2.6.6.6. DublinCore (http://dublincore.org/documents/dces)	158
2.7.Web semántica y Recuperación de la Información	162
2.7.1. Recuperación de la información: conceptualización	163
2.7.2. Sistemas de recuperación de la información (SRI) en la Web	164
2.7.3. SPARQL.....	171
2.7.4. Datos Enlazados	174
2.8.Referencias bibliográficas.....	177
CAPÍTULO III	190
3.1.CMS semánticos (S-CMS): análisis terminológico	191
3.2.Características.....	193

3.3.	S-CMS y gestión del contenido	194
3.3.1.	El contenido.....	194
3.3.2.	Estructura del S-CMS para la gestión de contenido semántico	197
3.4.	Procesos de organización y recuperación de la información desde los S-CMS	203
3.4.1.	Estándares y lenguajes de marcado más utilizados en los CMS semánticos	207
3.4.1.1.	XML: <i>Extensible Markup Language</i>	208
3.4.1.2.	RDF: <i>Resource Description Framework</i>	208
3.4.1.3.	OWL: <i>Web Ontology Language</i>	209
3.4.2.	S-CMS y ontologías	210
3.5.	Preservación de la información desde los S-CMS.....	211
3.6.	Referencias bibliográficas.....	224
CAPÍTULO IV	228
CAPÍTULO IV:	Marco metodológico de la investigación	229
4.1.	Contexto de investigación	229
4.2.	Objeto de investigación	232
4.3.	Definición y operacionalización de las variables.....	232
4.4.	Campo de investigación	237
4.5.	Tipo de investigación.....	237
4.6.	Enfoque de investigación	237
4.7.	Métodos de investigación	238
4.7.1.	Metodologías.....	238
4.7.2.	Métodos	239
4.7.3.	Instrumentos	241
4.7.4.	Herramientas	242
4.7.5.	Selección de la muestra.....	242

4.8.Etapas de investigación	244
4.9.Análisis de los recursos de información utilizados en la tesis.....	244
4.9.1. Los indicadores	244
4.9.2. Análisis métrico	246
4.10.Referencias bibliográficas	262
CAPÍTULO V.....	263
5.1.Introducción.....	264
5.2.Universidad de La Habana: Caracterización	265
5.3.Intranets en la UH.....	269
5.3.1. Universidad de La Habana	272
5.3.2. Facultad de Artes y Letras	275
5.3.3. Facultad de Biología	277
5.3.6. Facultad de Derecho	279
5.3.7. Facultad de Economía	280
5.3.8. Facultad de Turismo.....	281
5.3.9. Facultad de Filosofía e Historia	282
5.3.10. Facultad de Física	283
5.3.11. Facultad de Geografía	284
5.3.12. Facultad de Lenguas Extranjeras	286
5.3.13. Facultad de Matemática y Computación.....	287
5.3.14. Facultad de Psicología	289
5.3.15. Facultad de Química.....	289
5.3.16. Facultad de Español para no Hispanohablantes.....	290
5.3.17. Facultad de Educación a Distancia	292
5.3.18. CEPES.....	293
5.3.19. FLACSO.....	294

5.3.20.	CEAP.....	295
5.3.21.	CEEC	296
5.3.22.	BIOMAT.....	297
5.3.23.	CIEI	298
5.3.24.	CEDEM.....	299
5.3.25.	CIM	300
5.3.26.	IFAL.....	301
5.3.27.	Análisis del uso de los CMS para la construcción de Intranets.....	302
5.4.	Sistemas de gestión de información institucional automatizados de la UH	304
5.4.1.	Autenticador Único UH	305
5.4.2.	Consejo universitario	309
5.4.3.	Departamento jurídico.....	309
5.4.4.	Sistema de Relaciones Internacionales (SRI-UH)	310
5.4.5.	Sistema de Asignación de Responsables (Director)	311
5.4.6.	Sistema de Gestión para la Disponibilidad Financiera de la Universidad de La Habana (Disponibilidad Financiera)	311
5.4.7.	Sistema de gestión económica-financiera (EkoWeb).....	312
5.4.8.	Sistema de Gestión de Ingresos de la Dirección de Economía (Ingresos)	313
5.4.9.	Planificador UH.....	314
5.4.10.	Servicios Académicos Internacionales (OSAI-GP).....	315
5.4.11.	Sistema de Gestión de la Información de las investigaciones de la Universidad de La Habana (SGI-UH)	315
5.4.12.	Sistema de Gestión de la Nueva Universidad (SIGENU)	316
5.4.13.	Sistema de Gestión de Postgrado (SGP-UH)	317

5.4.14.	Sistema de Solicitud de Recursos (SSR)	318
5.4.15.	Sistema Análisis Periódico de Asistencia de la Universidad de La Habana APA-UH	319
5.4.16.	Análisis del uso de CMS en la construcción de sistemas de gestión automatizados	320
5.5. ..	Análisis de la calidad de los datos gestionados por los sistemas de gestión automatizados en la UH	322
5.5.1.	Diseño de la encuesta	322
5.5.2.	Selección de la muestra.....	323
5.5.3.	Resultados del Análisis de la calidad de los datos.....	324
5.6.	CMS más utilizados en la Universidad de La Habana	330
5.6.1.	Drupal.....	330
5.6.1.1.	Características.....	330
5.6.1.2.	Evaluación de características semánticas	332
5.6.2.	Joomla	339
5.6.2.1.	Características técnicas	340
5.6.2.2.	Evaluación características semánticas.....	342
5.7.	Referencias bibliográficas.....	344
CAPÍTULO VI		346
6.1.	Diseño de la propuesta de CMS Semántico para la UH: UH-WEB	348
6.1.1.	Principios y Requerimientos	348
6.1.2.	Objetivos.....	349
6.1.3.	Requisitos de Diseño.....	351
6.1.4.	Arquitectura.....	353
6.1.4.1.	Modelación lógica del CMS UH-Web	354
6.1.4.2.	Modelación estructural del CMS UH-Web	361

6.1.5.	Ontología y lenguajes ontológicos.....	364
6.1.5.1.	UH ontology	364
6.1.5.1.1.	Reutilización de ontologías	367
6.1.5.1.2.	Jerarquía de Clases	368
6.1.5.1.3.	Propiedades de los objetos	385
6.1.5.1.4.	Propiedades de los datos.....	392
6.1.5.1.5.	Anotaciones.....	394
6.1.5.1.6.	Relaciones de cardinalidad y ordenabilidad	396
6.1.5.1.7.	Gestión del conocimiento	397
6.1.5.2.	Lenguajes Ontológicos utilizados.....	399
6.1.6.	Esquema de Anotación Semántica	411
6.1.6.1.	Uso de Ontologías	412
6.1.6.2.	Anotea como Sistema de Anotación	414
6.1.6.3.	Infraestructura de Anotación Semántica.....	415
6.1.6.4.	Anotación Semántica Manual	423
6.1.7.	Recuperación de la Información.....	424
6.2.	Referencias bibliográficas.....	427
	CAPÍTULO VII.....	430
	Conclusiones.....	431
	Referencias bibliográficas	435
	CAPÍTULO VIII	436
	Trabajo Futuro	437
	Interoperabilidad y Datos enlazados	437
	Recuperación de Información	439
	Marcado Semántico como herramienta para la recuperación de la información	439

Integración de Herramientas de Publicación de Datos.....	441
Transformaciones modélicas, de formación y normativas de construcción: ...	442
Referencias bibliográficas	444
BIBLIOGRAFÍA CONSULTADA.....	446
Bibliografía Consultada.....	447
ANEXOS.....	468
Anexo 1: Encuesta 1	469
Anexo 2: Encuesta 2.....	473
Anexo 3: Comparación entre Joomla y Drupal	477
Anexo 4: Organigrama de La Universidad de La Habana	478

ÍNDICE DE TABLAS

Tabla.2.1. Formas de agrupar y clasificar información en RDF.....	104
Tabla 2.2. Reseña a las ontologías externas sobre las que hace referencia VIVO	152
Tabla 2.3. Elementos descriptivos de DC.....	159
Tabla.4.1. Presentación de la variable para la evaluación de las características semánticas de los CMS	231
Tabla.4.2. Presentación de la variable ORI.....	232
Tabla.4.3. Presentación de la variable GI.....	232
Tabla.4.4. Presentación de la variable BRI.....	234
Tabla.4.5. Resumen sobre la selección de la muestra	245
Tabla 4.6. Tipologías documentales utilizadas.....	244
Tabla 4.7. Cantidad de registros consultados por año	246
Tabla 4.8. Cantidad de registros consultados por palabras clave.....	248
Tabla 4.9. Autores con mayor cantidad de registros consultados.....	249

Tabla 4.10. Autores con mayor cantidad de registros citados.....	250
Tabla 5.1. Resultados de opinión sobre Drupal para el manejo de datos según los indicadores establecidos.....	321
Tabla 5.2. Resultados de opinión sobre Joomla para el manejo de datos según los indicadores establecidos	323
Tabla 5.3. Tabla comparativa entre Drupal y Joomla para el manejo de datos según los indicadores establecidos	324
Tabla 5.4. Resultado de la encuesta realizada a especialistas sobre Drupal...	333
Tabla 5.5. Resultado de la encuesta realizada a especialistas sobre Joomla ..	338

ÍNDICE DE FIGURAS

Fig.1.1. Esquema de Bramscher y Butler	45
Fig.1.2. Ejemplo del comportamiento de la frecuencia de publicaciones sobre CMS hasta el año 2015	46
Fig.1.3. Esquema sobre evolución de las páginas Web y los lenguajes que las sustentan	48
Fig.1.4. Esquema evolutivo de los CMS y su relación con el desarrollo de lenguajes de programación para la Web	52
Fig. 1.5. Estructura del CMS según Boiko	53
Fig. 1.6. Estructura del CMS según McKeever	54
Fig. 1.7. Estructura del CMS según Robertson	55
Fig.1.8. Estructura del CMS según Michelinakis	56
Fig.1.9. Estructura de los CMS según la interpretación dada a las ideas de Jiménez y Moreo	57
Fig.1.10. Estructura de los CMS según Sánchez (2011)	58
Fig.1.11. Clasificación de los CMS desde la dimensión jurídica	65
Fig.1.12. Clasificación de los CMS desde la dimensión de la semántica documental.....	68
Fig.1.13. Esquema resumen sobre la clasificación de los CMS	69
Fig.1.14. Interdisciplinariedad sistémica integrada de la Ciencia de la Información	70

Fig.1.15. Interdisciplinariedad lineal de la Ciencia de la Información	71
Fig.2.1.Ejemplo de implementación de Web Semántica: Visualización de la Página principal de Europeana.....	79
Fig.2.2.Mapa conceptual del término Web Semántica	80
Fig. 2.3. Implicaciones del término semántica.	82
Fig. 2.4. Evolución de la Web	84
Fig. 2.5. Diferencia gráfica entre la Web sintáctica y la Web semántica	87
Fig. 2.6. Evolución de la Web: DESDE LA Web 1.0 hasta la Web 7.0	89
Fig. 2.7. Modelización de la transición de la web actual a la web semántica ...	90
Fig. 2.8. Base semántica de la Web semántica	91
Fig.2.9. Modelo de estructura en capas de la Web semántica	92
Fig.2.10. Especificaciones y niveles conceptuales de abstracción de las diferentes tecnologías de la Web Semántica.....	94
Fig.2.11.Componentes de XML según W3C	99
Fig.2.12.Componentes de RDF: equivalencias lógico-lingüísticas.....	102
Fig.2.13. Tripletas RDF: ejemplificación de las equivalencias lógico-lingüísticas	103
Fig. 2.14. Constructores RDF y RDF Schema	105
Fig. 2.15. Propiedades de OWL.....	110
Fig. 2.16. Constructores OWL Lite, DL y Full	107
Fig.2.17. Propuesta de estructura de una ontología del contexto universitario	119
Fig.2.18. Evolución de los diferentes tipos de sistemas organización del conocimiento	120
Fig. 2.19. Procedimiento para el diseño y evaluación de ontologías.....	124
Fig.2.20. Propuesta de SENSUS para vincular términos de dominio a una ontología de amplia cobertura	126
Fig. 2.21. Fases de la metodología ONTOLOGY DEVELOPMENT 101.....	127
Fig 2.22. Metadatos identificativos de SKOS.....	137
Fig. 2.23. Evolución de las versiones de SKOS	137
Fig. 2.24. Clases de la ontología SKOS	138
Fig.2.25. Relación de SKOS con otros vocabularios	142

Fig 2.26. Metadatos identificativos de BIBO	145
Fig. 2.27. Evolución de las versiones de BIBO	145
Fig. 2.28. Relación de BIBO con otros vocabularios	146
Fig. 2.29. Clases de la ontología BIBO	147
Fig. 2.30. Metadatos identificativos de VIVO	148
Fig. 2.31. Evolución de las versiones de VIVO	149
Fig. 2.32. Relación de VIVO con otros lenguajes	150
Fig.2.33.Ontologías integradas en VIVO Ontology y sus clases principales ...	151
Fig. 2.34. Metadatos identificativos de FOAF	153
Fig. 2.35. Evolución de las versiones de FOAF	154
Fig. 2.36. Vocabulario FOAF	154
Fig. 2.37. Relación de FOAF con otros vocabularios	155
Fig. 2.38. Metadatos identificativos de Geopolitical ontology	156
Fig. 2.39. Relación de Geopolitical ontology con otros vocabularios	157
Fig. 2.40. Metadatos identificativos de Dublin Core	158
Fig.2.41. Evolución de las versiones de DC	158
Fig. 2.42. Relación de DC con otros vocabularios	162
Fig.2.43. Eficacia de la búsqueda en texto libre en relación a la evolución de la Web	165
Fig.2.44. Relación entre el incremento de la capacidad de búsqueda y el uso de metadatos	166
Fig.2.45. Vertiente semántica y Pragmática de una ontología.	168
Fig.2.46. Uso del Where y el Select en las Consultas SPARQL	172
Fig.2.47. Ejemplo de Consultas SPARQL	172
Fig.2.48. Representación gráfica de la nube de datos enlazados	175
Fig.3.1. Implicaciones del término “contenido” para el estudio de los CMS ...	196
Fig.3.2.Representación gráfica de la estructura genérica de capas de un S-CMS	197
Fig.3.3.Estructura compleja de capas de un S-CMS.....	199
Fig.3.4.Estructura del S-CMS FLERSA.....	201
Fig.3.5.Dimensiones de análisis de la preservación de la información digital	211

Fig.3.6.Asignación de metadatos a los contenidos	215
Fig.3.7. Ejemplo de control de versiones.	219
Fig. 4.1: Tipologías de productos Web construidos en la UH.....	228
Fig. 4.2: Balance sobre los tipos documentales ¡Error! Marcador no definido.	245
Fig. 4.3: Tipos documentales utilizados	246
Fig. 4.4: Balance sobre los niveles de actualización de la bibliografía consultada	247
Fig. 4.5: Relación entre la cantidad de documentos consultados y los citados de los principales autores trabajados.....	253
Fig. 4.6: Colaboración autoral de la bibliografía consultada.....	258
Fig. 5.1. Carreras que se estudian en el curso regular diurno (2015).....	261
Fig. 5.2. Carreras que se estudian por encuentro (2015)	263
Fig. 5.3. Carreras que se estudian en continuidad de estudio (2015).....	263
Fig. 5.4. Carreras que se estudian a distancia (2015).....	263
Fig.5.5. Distribución geografica de las diferentes estructuras de la UH	264
Fig. 5.6. Organigrama de la Universidad de La Habana.....	265
Fig.5.7. CMS identificados en la UH (intranets) en estudio exploratorio de 2009	266
Fig.5. 8. Existencia de Intranets en la UH, enero/ 2016.....	267
Fig.5. 9. Página de inicio de la Intranet UH (enero/2009)	268
Fig.5.10. Página de inicio de la Intranet UH (enero/2016).....	269
Fig.5.11. Diferentes áreas de la Página de inicio de la Intranet UH.....	260
Fig.5.13. Página de inicio de la intranet de la facultad de Biología.....	271
Fig.5.14. Página de inicio de la intranet de la Facultad de Comunicación.....	272
Fig.5.15. Página de inicio de la intranet de la Facultad de Contabilidad y Finanzas.....	273
Fig.5.16. Página de inicio de la intranet de la Facultad de Derecho.....	274
Fig.5.17. Página de inicio de la intranet de la Facultad de Economía.....	275
Fig.5.18. Página de inicio de la intranet de la Facultad de Turismo.....	276
Fig.5.19. Página de inicio de la intranet de la Facultad de Filosofía e Historia	277
Fig.5.20. Página de inicio de la intranet de la Facultad de Física	278

Fig.5.21. página de inicio de la intranet de la Facultad de Geografía.....	279
Fig.5.22. Página de inicio de la intranet de la Facultad de Lenguas Extranjeras	280
Fig.5.23. Página de inicio de la intranet de la Facultad de Matemática	281
Fig.5.24. Página de inicio de la intranet de la Facultad de Psicología	283
Fig.5.25. Página de inicio de la intranet de la Facultad de Química.....	284
Fig.5.26. Página de inicio de la intranet de FENHI	285
Fig.5.27. Página de inicio de la intranet de la Facultad de Educación a Distancia	286
Fig.5.28. Página de inicio de la intranet del CEPES.....	287
Fig.5.29. Página de inicio de la intranet del FLACSO.....	288
Fig.5.30. Página de inicio de la intranet del CEAP.....	289
Fig.5.31. Página de inicio de la intranet del CEEC	290
Fig.5.32. Página de inicio de la intranet de BIOMAT	291
Fig.5.33. Página de inicio de la intranet del CIEI	292
Fig.5.34. Página de inicio de la intranet del CEDEM.....	293
Fig.5.35. Página de inicio de la intranet del CIM	294
Fig.5.36. Página de inicio de la intranet IFAL.....	295
Fig.5.37. Intranet de la UH y proporción de Software utilizados, enero/ 2016	296
Fig.5.38. CMS identificados en las Intranets de la UH.....	296
Fig. 5.39. Esquema de comunicación entre los sistemas de gestión en la UH	298
Fig. 5. 40. Visualización de la página de inicio de acceso al sistema Autenticador Único UH o Directorio Único UH	299
Fig. 5. 41. Comunicación de los sistemas UH con el directorio Único	299
Fig. 5. 42. Ejemplo 1 de registro de un usuario en el Directorio Único: Datos personales.....	300
Fig. 5. 43. Ejemplo 2 de registro de un usuario en el Directorio Único: Datos docentes	300
Fig. 5. 44. Directorio Único: búsqueda de trabajadores	301
Fig. 5. 45. Directorio Único: búsqueda de estudiantes	301
Fig.5.46. Página de inicio del sitio del Consejo Universitarios de la UH	302
Fig.5.47. Página de inicio del sitio Departamento Jurídico	303

Fig.5.48. Página de inicio del sitio SRI-UH	304
Fig.5.49. Página de inicio de Director	305
Fig.5.50. Página de inicio del sitio de gestión de Disponibilidad Financiera ...	305
Fig.5.51. Página de inicio del sitio Eko-Web	306
Fig.5.52. Página de inicio para la gestión on-line	307
Fig.5.53. Página de inicio de Ingresos	307
Fig.5.54. Página de inicio para entrar al Planificador	308
Fig. 5.55. Página de trabajo.....	308
Fig.5.56. página de inicio de OSAI-GP	309
Fig.5.57. Página de inicio del SGI-UH.....	310
Fig.5.58. Página de inicio del SIGENU	311
Fig.5.59. Página de inicio al SGO-UH	312
Fig.5.60. Página de inicio de SSR.....	313
Fig.5.61. Solicitud de un local: reservación en línea	313
Fig.5.62. Página de inicio del sitio APA-UH	314
Fig. 5.63. Uso de CMS en la construcción de sistemas de gestión automatizados	317
Fig.5.64. Indicadores para la valoración de calidad de datos	319
Fig. 5.65. Composición educacional del Consejo Universitario de la UH.....	321
Fig. 5.66. Comportamiento de los indicadores recogidos.....	325
Fig. 5.67.ModeloRDF en Drupal 7: Representación de nodos, usuarios, roles, temas y versiones con RDF	329
Fig. 5.68.Módulo básico del soporte RDF en Drupal 7	330
Fig. 5.69.Módulos que amplían el soporte nativoRDF en Drupal 7	330
Fig. 5.70.Vocabularios RDF externos en Drupal 7: Incorporación de vocabularios mediante External RDF VocabularyImporter.....	331
Fig. 5.71.Tipos de Contenidos y RDF en Drupal 7: Definición de objetos RDF a partir de tipos de contenidos en Drupal	332
Fig. 5.72. Soporte RDFa en Drupal 7: Inserción de RDFa en el código XHTML332	
Fig. 5.73.Soporte SPARQL en Drupal 7: El ecosistema SPARQL en Drupal....	333
Fig.6.1. Modelo lógico del CMS: UH-WEB	348

Fig. 6.2. Enlazado de datos desde Silk-LSL	353
Fig. 6.3. Transformación de datos desde Silk-LSL	353
Fig.6.4. Fase de Uso de datos semánticos por parte del usuario	355
Fig.6.5. Modelo estructural del CMS: UH-WEB (elaboración propia).....	357
Fig.6.6. UH ontology: Conteo de las estructuras principales	361
Fig.6.7. UH ontology: Axiomas.....	361
Fig.6.8. UH ontology: Objetos	361
Fig.6.9. UH ontology: Datos	362
Fig.6.10. UH ontology: Personas	362
Fig.6.11. UH ontology: Anotaciones	362
Fig.6.12. UH ontology: Clases principales.	363
Fig.6.13. UH ontology: Clase Agent y su relación con otras clases y subclases de la ontología (1).....	365
Fig.6.14. UH ontology: Clase Agent y su relación con otras clases y subclases de la ontología (2).....	366
Fig.6.15. UH ontology: Relaciones de subordinación de la subclase Organization	366
Fig.6.16. UH ontology: subclase Organization y relación con otras clases y subclases.	367
Fig.6.17. UH ontology: Relaciones de la subclase Person	368
Fig.6.18. UH ontology: subclase Person y su relación con otras clases y subclases.	369
Fig.6.19. UH ontology: Clases Collection, InformationResource, Document y Resource	369
Fig.6.20. UH ontology: clase Collection	370
Fig.6.21. UH ontology: clase Collection y relaciones entre sus subclases.....	370
Fig.6.22. UH ontology: Relaciones de la clase Collection	370
Fig.6.23. UH ontology: Subclases definidas dentro de Document	373
Fig.6.24. UH ontology: Relaciones de la clase Document	374
Fig.6.25. UH ontology: Relaciones de la clase Document	375
Fig.6.26. UH ontology: Relaciones de la clase Resource.....	377
Fig.6.27. UH ontology: Relaciones de la clase Concept	378

Fig.6.28. UH ontology: Relaciones de la clase Credential	378
Fig.6.29. UH ontology: Relaciones de la clase EducationalTraining	378
Fig.6.30. UH ontology: Clase Event.....	377
Fig.6.31. UH ontology: Relaciones de jerarquía de la clase Event.....	378
Fig.6.32. UH ontology: Relaciones de la clase Project.....	378
Fig.6.33. UH ontology: Relaciones de la clase Software	379
Fig.6.34. UH ontology: Relaciones de la clase Area	379
Fig.6.35. UH ontology: Sistema de anotación.....	388
Fig.6.36. UH ontology: Anotaciones de la ontología.....	389
Fig.6.37. UH ontology: Ejemplo de relaciones de cardinalidad desde la Clase Document.....	390
Fig. 6.38. Gestión del conocimiento desde UH ontology	391
Fig.6.39. Clases fundamentales de SKOS.....	393
Fig.6.40. Vista HTML del uso de BIBO en UH ontology	394
Fig.6.41. Vista HTML de UH Ontology. Ejemplificación de VIVO.....	395
Fig. 6.42. Vista HTML de UH Ontology. Ejemplificación de VIVO y su interacción con las clases de otros vocabularios ontológicos	396
Fig. 6.43. Clase Event y manejo de sus propiedades	397
Fig. 6.44. Visualización de la clase Event desde la ontología. Vista HTML	398
Fig.6.45.Modelo de la clase Event	398
Fig6.46. Clase Area y manejo de sus propiedades	399
Fig.6.47. Enfoque de integración semántica de datos geoespaciales.....	400
Fig. 6.48. Clase Agente y manejo de sus propiedades	404
Fig.6.49. Subclase LearningResourceType y manejo de sus propiedades	405
Fig.6.50. Taxonomía de clases de infraestructura de Annotea.....	409
Fig.6.51. Taxonomía de clases de infraestructura del CMS UH WEB.....	410
Fig.6.52. Instancia de la clase AnnotationID (elaboración propia)	415
Fig. 6.53. Consulta SPARQL para UH Ontology	418

INTRODUCCIÓN

Introducción

La Ciencia de la Información (CI) es una disciplina relativamente joven, surgida de la confluencia de varios factores. Entre ellos, la manifestación de un conjunto trascendental para el desarrollo de la sociedad, las Nuevas Tecnologías de la Información y las Comunicaciones (NTIC). Estas generaron una revolución en los modos de hacer y pensar las ciencias lo cual influyó significativamente en la necesidad de organizar y gestionar un enorme flujo de información. A partir de la década del ochenta del siglo XX, la información es abordada como un recurso estratégico y vital, capaz de acelerar el progreso en todos los espacios: sociales, económicos, tecnológicos, políticos, militares, etc.

La aparición en 1991 de la Word Wide Web, más conocida como WWW y creada por Tim Berners Lee, modificó la delimitación del tiempo y el espacio de la información, pues propuso un nuevo sistema de "hipertexto" para compartir documentos. Ante el nuevo contexto, Albuquerque proyecta la necesidad de gerenciar y controlar ese gran volumen de información, **"almacenarla y caracterizar su contenido, priorizar su uso – de acuerdo con las diferentes comunidades informacionales – y promover una difusión selectiva y retrospectiva"** (Albuquerque, 1998). La interacción en tiempo real en la red se materializa, a decir de este autor, en una ubicuidad del texto y de la información en sentido general.

La sociedad actual se caracteriza por el interés para aprovechar intensamente este recurso, en todas las esferas. Es poco común no percibir en las organizaciones la tendencia, de sus distintos miembros y a diferentes niveles, al uso adecuado de la información en función del cumplimiento de su misión y los objetivos trazados. Por lo general, la confección de productos como una Intranet, un portal, una biblioteca digital, tributan a ello y se requiere de un estudio profundo de toda la información que circula dentro de la institución, para luego realizar su diseño, programación e implementación. Como consecuencia el desarrollo de la actividad de organización y recuperación de información demande mejoras en los procesos de entendimiento entre plataformas de contenido.

Desde que se gestó la web como sistema de información, grandes han sido los saltos cualitativos que ha experimentado la gestión de información a partir de la

incorporación en su diseño de nuevas tecnologías que facilitan su transmisión, organización y recuperación eficiente para los usuarios. Aportes importantes en el desarrollo de instrumentos de organización conceptual y terminológica como los tesauros, taxonomías y ontologías, han sido generados y/o implementados desde diferentes dominios profesionales.

La llegada del concepto de Web semántica, trajo una nueva revolución en el terreno de la web, los usos de ontologías y los marcadores semánticos hicieron que estas acogieran una filosofía que, si bien no desvirtuaba las formas tradicionales de navegación, dotaba a la web (sintáctica) de nuevas formas de razonamiento cada vez más apegadas a las necesidades de sus usuarios finales, al expresar la forma de comunicación de un dominio y sus particularidades léxicas.

Paralelamente, dentro del campo de creación de productos digitales en la Web, los Sistemas Gestores de Contenido (CMS: siglas en inglés de Content Management System) han generado un impacto significativo, motivado por el manejo de manera independiente de diseño y contenido y otras facilidades para la gestión de la información dentro del sitio. Además originaron rebaja en los costos del proceso de elaboración e implementación de los productos electrónicos para la Web, dados esencialmente por una disminución del tiempo invertido, la socialización y descentralización de los accesos para la actualización de la información y el aumento de las **potencialidades desarrolladas. Ellos "representan el paradigma actual en el diseño y desarrollo de sistemas y servicios de información basados en la Web"** (Pastor Sánchez and Martínez Méndez, 2009).

Sin embargo, no todos los CMS cuentan con disposiciones para el desarrollo de los procesos de organización y recuperación de la información con herramientas semánticas, lo que problematiza el intercambio y recuperación de la información y se frena la interoperabilidad entre plataformas.

A esta situación socio-histórica y tecnológica, no están ajenas las universidades cubanas desde donde se reconocen las dificultades semánticas que enfrentan en la construcción de sus productos web institucionales.

Escenario de investigación

En Cuba, alrededor de los años noventa, con la generalización de los lineamientos estratégicos para la informatización, comienza un incremento gradual en el uso de las NTIC. La aplicación del Programa de Informatización de la Sociedad, que sirve como plataforma de ejecución de dichos lineamientos, tributó al aumento paulatino de equipamiento tecnológico en las universidades, la enseñanza y aplicación de programas y herramientas computacionales a todos los niveles educacionales y el uso de las NTIC como apoyo a la docencia.

La implementación de las NTIC en función de la instrucción universitaria propició procesos de aprendizaje con la finalidad de orientar al estudiante hacia la creación de su propio conocimiento, a partir del conjunto de recursos e informaciones disponibles. Para acometer esta nueva tarea el profesor ha necesitado trabajar conjuntamente con diversos especialistas (informáticos, pedagogos, profesionales de la información, diseñadores, etc.); lo que exige del trabajo cooperativo, pues la simple presencia de una computadora conectada a Internet o en red no garantiza un adecuado aprovechamiento de las tecnologías, pese a que sus posibilidades sean infinitas. Las NTIC no generan automáticamente nuevos saberes. La herramienta ha de estar al servicio de lo educativo y no a la inversa.

En el espacio universitario se crean diferentes instrumentos con el objetivo de mejorar la gestión de información dentro de la institución. Los sitios web, portales, intranet y todo tipo de productos digitales, cobran auge dentro de los diferentes Centros de Educación Superior (CES) y se sitúa entre los puntos de seguimiento y control desde los Ministerios a los cuales dichos centros se subordinan.

En el 1998, dos universidades en la capital cubana contaban con página web. Alrededor del año 2000, algunas instituciones introdujeron los CMS como herramientas de trabajo en la creación de sitios temáticos. Entre los centros considerados pioneros en estas prácticas se encontraban universidades de la capital como la **Universidad de La Habana (UH)** y el **Instituto Superior Politécnico "José Antonio Echevarría" (CUJAE)**.

Entre el 2000 y 2004 aumentó significativamente el número de CES con páginas institucionales o intranet. Inicialmente, en la mayoría de ellas, se trabaja con el uso

de software como Macromedia Dreamweaver, programación manual de las páginas a partir del manejo de los códigos en lenguaje HTML, pequeños programas para enriquecer el diseño y funcionalidad como Microsoft FrontPage, etc. Se caracterizaban por ser páginas estáticas, meramente informativas.

A partir de 2008, con el desarrollo del tema, tanto en experiencias teóricas como prácticas, aparecen de manera explícita en varios de los documentos estratégicos de planificación en los CES, la necesidad de compartir información y trazar estrategias metodológicamente coordinadas. El uso de los CMS de tipo software libre desde las universidades se introduce y desarrolla dentro de este contexto, favoreciéndolo a su vez, dadas las potencialidades en su acceso, aprendizaje y uso.

Según los resultados de un estudio exploratorio realizado por esta autora en el año **2009, “el impacto de los CMS en la construcción de sitios Web e Intranet de centros de educación superior enmarcados en la capital, ha sido positivo e importante para elevar la calidad de los productos Web implementados en las universidades”** (Rosell León, 2009). A partir de dicho estudio se pudo conocer qué, entonces, el 87.5% de las universidades incluidas en el estudio tenían productos electrónicos implementados mediante el uso de los CMS, con una amplia variedad de los mismos (Plone, Drupal, Joomla, PHP-Nuke y Php-fusion; aparecen en ese orden de preferencia como los más utilizados). Años más tarde, un sondeo genérico arrojó un aumento al 93% en el uso de CMS, siendo Drupal y Joomla los CMS mayormente manejados (Rosell León, 2009, Rosell León, 2011).

La UH, líder junto a otros CES del país en estos temas (como la CUJAE, la UCI y la UCLV), cuenta con un gran número de productos Web, ya no solo intranets y páginas Web institucionales. Se generan bibliotecas digitales, repositorios, wikis, etc. En su mayoría construidos en Drupal y Joomla. Sin embargo, refieren limitaciones en la interoperabilidad de contenidos, manejo y gestión de semánticas para la organización y recuperación de la información, lo que incide directamente en la satisfacción de estudiantes, profesores, investigadores y trabajadores en general de la universidad.

La certeza de estas afirmaciones nos condujo a proponer las bases teóricas para desarrollar un estudio más integral, con mayor alcance y sustento teórico.

Problema de investigación

En la UH se construyen las páginas web institucionales (a nivel central como universidad y de manera particular de cada una de las facultades y centros que la componen) con la utilización de CMS, esencialmente Drupal y Joomla.

Como universidad, se ha proyectado para que sus productos web se integren de manera eficiente en la dinámica de la web 2.0 con el fin de fortalecer la visibilidad de sus contenidos. Aunque los CMS utilizados permiten crear y mantener los sitios **web con facilidad, “las tecnologías emergentes de la web semántica (XML, RDF y OWL) aportan descripciones explícitas de los recursos de la Web (...) con objeto de facilitar la interpretación de los documentos y de realizar procesos inteligentes de captura y tratamiento de información (para permitir la interoperabilidad)”** (Navarro Galindo, 2012).

La inexistencia hoy de un CMS semántico en la UH dificulta la coherencia entre los procesos de gestión y recuperación de información actuales y las políticas trazadas en pos del desarrollo del sitio Web, por las limitaciones en cuanto a formato, integración y recuperación de los contenidos existentes. Sus sitios coexisten aislados entre sí y no es posible la unificación y reutilización de la información entre ellos.

Pregunta de investigación

¿Cómo facilitar la organización, gestión y recuperación de la información en los sitios Web de la Universidad de La Habana?

Sistema de objetivos

Objetivo general:

Diseñar un modelo de CMS semántico para los sitios web de la Universidad de La Habana

Objetivos específicos:

- Determinar los referentes teóricos – metodológicos que sustentan el diseño de un CMS semántico.
- Caracterizar a la Universidad de La Habana como marco aplicado de la investigación.

- Evaluar los CMS implementados en la UH.
- Proponer modelo de CMS semántico para la UH.

Se deslinda entonces como objeto de estudio la *modelación de un CMS semántico* y como campo de acción la *modelación del S-CMS UH-WEB para la UH*.

Antecedentes

En la Universidad de La Habana, particularmente dentro de la especialidad de Bibliotecología y Ciencia de la Información (BCI), se han desarrollado algunos trabajos de investigación bajo el tema de los CMS, razón por la cual constituyen antecedente para la investigación. Ellos son:

- Trabajo de Diploma de la Lic. Mabel Rodríguez Mederos: La gestión de información y el software libre. Herramientas para automatizar un centro de información (2003).
- Trabajo de Diploma del Lic. David González Romero: CMS: Software libre para implementar la intranet del Centro Nacional de Áreas Protegidas (2006).
- Tesis de Maestría de la MsC. Yorbelis Rosell León: Impacto de los Sistemas Gestores de Contenido (CMS) en Centros de Educación Superior de Ciudad de La Habana (2009) defendida como Diploma de Estudios Avanzados dentro del Doctorado en Documentación e Información Científica (de la Universidad de Granada para impartir en La Universidad de La Habana).

A nivel internacional, se conoce la existencia de estudios similares, como:

- Tesis doctoral de José Luis Navarro Galindo: *FLERSA: un sistema semántico de gestión de contenido web (S-CMS)*, 2012. Universidad de Granada. Defendida como parte del Programa de Doctorado sobre Métodos y Técnicas Avanzadas del Desarrollo de Software

Alcance y novedad de los resultados

La tesis presenta un modelo personalizado para la UH desde donde se enriquece el concepto de CMS con las herramientas de la Web semántica. Desde un punto de vista práctico, la propuesta tributa a la construcción de sitios personalizados, con mejor organización de los contenidos gestionados en la universidad y por tanto,

incide directamente en los procesos de búsqueda y recuperación de la información de las intranets y sitios construidos desde la UH.

Además, como documento científico, resume importantes referentes teóricos y metodológicos sobre temas cruciales para la Ciencia de la Información, al reflejar temáticas referidas a la Web semántica, los CMS, ontologías, tratamiento de datos, entre otros. Los mismos servirán como una visión de la realidad que puede extrapolarse, en diferentes niveles de aplicación y en diversas esferas y contextos.

La investigación permitirá a las autoridades y factores que dirigen el proceso de informatización, así como a los investigadores e interesados en el tema, las herramientas teóricas y metodología para forjarse una visión integral a posteriori como criterio de evaluación y la toma de decisiones.

En el estudio se asumen los roles del Profesional de la Información como miembro pleno del grupo multidisciplinario que se implica activamente en la construcción de productos web dentro de las universidades. Este rol se asemeja con el del arquitecto de la información web, término que en Cuba nos es empleado o particularmente reconocido en la praxis nacional. Por dicha razón, se prefiere emplear el término profesional de la información, al ser más genérico.

Se estudia a los CMS desde la óptica de los procesos de producción de información.

Limitaciones del estudio

Una mirada como la que exige esta investigación depende especialmente de una formación continua y crítica a la que no se llega sólo por convencimiento; la pretendida capacidad para distinguir lo esencial muchas veces entra en contradicción con una apoyatura múltiple y compleja que más de una vez inquirió en esta autora si había escogido acertadamente los argumentos.

Una investigación en la que se pretenda modelar un CMS semántico, por la complejidad y alcance que tendría, requeriría de un tiempo mayor que el invertido en el presente estudio, para la implementación y evaluación de la puesta en práctica.

Estructura Capítular

La tesis se compone de ocho capítulos:

- ✓ Capítulo 1: MARCO TEÓRICO REFERENCIAL: SISTEMAS GESTORES DE CONTENIDO (CMS)

Se realiza un análisis terminológico conceptual sobre los CMS. Se aborda su génesis, devenir histórico, funcionalidades, tipologías. Así como la visión que de ellos se tiene desde el contexto de las Ciencias de la Información (relaciones y aportes bidireccionales).

- ✓ Capítulo 2: MARCO TEÓRICO REFERENCIAL: WEB SEMÁNTICA

Se realiza un análisis terminológico, génesis, estructura y aspectos distintivos de la recuperación de la información de la web semántica. Se aborda además su relación con estándares y lenguajes de marcado y ontologías.

- ✓ Capítulo 3: MARCO TEÓRICO REFERENCIAL: CMS SEMÁNTICO (S-CMS)

Se estudia conceptualmente el término de S-CMS, su evolución, estructura y modelación. Se abordan otros aspectos que colateralmente inciden en su implementación y desarrollo como: estándares, la anotación semántica, métodos de marcado semántico, datos enlazados.

- ✓ Capítulo 4: MARCO METODOLÓGICO DE LA INVESTIGACIÓN

Se analiza contexto, objeto y métodos de investigación, técnicas para la recopilación de información, se describen las etapas del estudio. Finalmente se presenta un estudio bibliométrico de los recursos de información utilizados en la investigación.

- ✓ Capítulo 5: MARCO APLICADO DE LA INVESTIGACIÓN

De manera particular se estudian los CMS más utilizados en la UH. Se describen las características y estructura general de la Universidad, así como las páginas construidas con CMS (intranets y sistemas de gestión de

información institucional). Se presentan los parámetros de evaluación y los resultados obtenidos

✓ Capítulo 6: PROPUESTA MODÉLICA DEL CMS SEMÁNTICO UH-WEB

Se realiza una presentación de la propuesta de CMS semánticos para la UH. Se describen los requisitos de diseño, arquitectura y los requerimientos de anotación semántica, así como su ontología núcleo.

✓ Capítulo 7: CONCLUSIONES

Se presentan conclusiones sobre los contenidos teóricos abordados y la propuesta de CMS UH-Web.

✓ Capítulo 8: TRABAJO FUTURO

Se presentan algunas líneas generales que conducirían al desarrollo del modelo desde una perspectiva de futuro.

Referencias bibliográficas

- ALBUQUERQUE, A. 1998. Los tres mundos de la Ciencia de la Información. *Ciencias de la Información*, 29 (3).
- NAVARRO GALINDO, J. L. 2012. *FLERSA: un sistema semántico de gestión de contenido web (S-CMS)*. Tesis doctoral, Universidad de Granada.
- PASTOR SÁNCHEZ, J. A. & MARTÍNEZ MÉNDEZ, F. J. 2009. Aplicación de tesauros, taxonomías y ontologías en los sistemas de gestión de contenidos mediante tecnologías de la Web Semántica. *Ibersid*, 143-153.
- ROSELL LEÓN, Y. 2009. *Impacto de los Sistemas Gestores de Contenido (CMS) en Centros de Educación Superior de Ciudad de La Habana*. Diploma de Estudios Avanzados del Doctorado en Documentación e Información Científica, Universidad de La Habana (Cuba) – Universidad de Granada (España).
- ROSELL LEÓN, Y. 2011. Los CMS y su impacto en Centros de Educación Superior: estudio exploratorio. In: *IV Taller "La Virtualización de la Educación Superior"*, Evento Provincial UNIVERSIDAD 2012, Cuba.

CAPÍTULO I

LOS SISTEMAS GESTORES DE CONTENIDOS (CMS)

CAPÍTULO 1: Marco teórico referencial: Los Sistemas Gestores de Contenido (CMS)

En el presente capítulo se realiza un análisis terminológico conceptual sobre los CMS. Se aborda su génesis, devenir histórico, funcionalidades, tipologías. Así como la visión que de ellos se tiene desde el contexto de las Ciencias de la Información (relaciones y aportes bidireccionales).

1.1. Sistemas de Gestión de Contenidos: análisis terminológico

El término Sistemas Gestores de Contenido, por lo general, se identifica en la literatura por las siglas CMS, que en inglés responden a la abreviatura de *Content Management System*.

“El concepto de Sistemas de Gestión de Contenidos era desconocido hasta principios de los años noventa. No obstante, algunas de sus funciones ya se realizaban con aplicaciones independientes tales como editores de texto y de imágenes, bases de datos y programación a medida” (Osuna Alarcón and De la Cruz Gómez, 2010). Originalmente fueron usados para la publicación de sitios Web. Con el tiempo han evolucionado y ampliado sus especializaciones, así como el alcance del significado del término.

A partir del contenido semántico del vocablo, es posible suponer que es la máquina o el sistema informático lo que gestiona los contenidos de manera independiente, sin la intervención humana. Por tal razón, algunos autores prefieren nominalizarlo como "sistema de soporte a la gestión de contenidos", pues son los modos de comunicación los que llevan a gestionar contenidos de manera coherente.

Begoña, Bazán y Hungaro plantean que puede tornarse compleja la tarea de definir estas herramientas por la variedad de funcionalidades requeridas. Son del criterio que esto ha llevado a asumir el término “como un concepto más que una tecnología o producto (...) que abarca un conjunto de procesos que facilitan la creación y administración de contenido colaborativo” (Begoña Rodríguez et al., 2008).

Sin embargo, como criterio recurrente entre los especialistas aventurados a definir conceptualmente el fenómeno de los CMS, confluyen los enfoques dirigidos a las funcionalidades que deben cumplir como software.

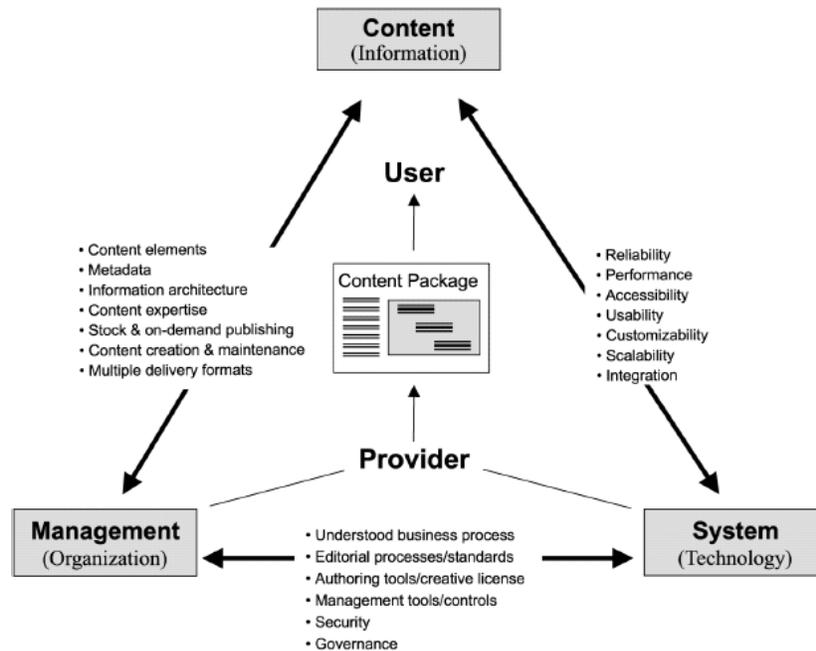
A continuación se presentan, cronológicamente, un grupo de conceptos con la intención de analizar los elementos aportados de mayor significación.

Boiko es uno de los autores más citados en los trabajos que abordan la conceptualización de los CMS. Dicho especialista define a los CMS como **“un sistema que recopila, administra y publica la información y funcionalidad... eso es en parte hardware y software, procesos en parte y en parte un vehículo de la organización”** (Boiko, 2001). Como elemento importante se reconoce al CMS como un sistema, término que se asume implícita y/o explícitamente en todos los conceptos.

En la definición aportada por Robertson se introduce como parte de la definición la capacidad de trabajar de manera individualizada el contenido del diseño visual de los productos Web (Robertson, 2003). Dicho autor señala que un CMS apoya la creación, gestión, distribución, publicación y divulgación de la información corporativa. Integra de manera completa el ciclo de vida de las páginas de sus sitios, al proporcionar herramientas sencillas para crear el contenido, a través de la publicación, y finalmente archivando. Además, proporciona la capacidad de manejar la estructura del sitio, la apariencia de las páginas publicadas, y la navegación orientada o en función del usuario.

Michelinakis **señala que** **“los Sistemas de Gestión de Contenido no son sólo un producto o una tecnología. CMS es definido como un término genérico que se refiere a un amplio rango de procesos que sostienen la “siguiente generación”** de los sitios Web medianos y grandes. La gestión de contenido es un proceso que se encarga de la creación, almacenamiento, modificación, recuperación y presentación de datos o contenido” (Michelinakis, 2004). En este concepto se reconoce la importancia de la llegada de los CMS en el mundo de la construcción de las páginas web y por transitividad en Internet, así como las formas de interacción de los usuarios y gestión de la información.

Bramscher y Butler, definen al CMS desde la explicación del esquema presentado en la figura 1.1 (Bramscher and Butler, 2006).



Source: Adapted from White (2005, p. 11)

Fig. 1.1. Esquema de Bramscher y Butler (Bramscher and Butler, 2006)

Según este autor, el CMS se define a partir de la relación implícita entre las estructuras, contenido (información), gestión (organización), sistema (la tecnología). Defiende que el CMS debe ser visto como un conjunto coordinado de las tecnologías y los procesos. Las funciones críticas para el éxito de un CMS como tecnología son: posibilidad de personalizar el material digital u objetos (contenido) por personal no técnico; socializar las decisiones relativas a la arquitectura del sitio y la distribución de la autoría y funciones de edición (gestión) y proporcionar los mecanismos para llevar esto a cabo (el sistema). Ellas permiten a una organización satisfacer las necesidades de sus usuarios (Bramscher and Butler, 2006).

Para Ortegón, un CMS "es un sistema que permite actualizar el contenido de una página Web sin involucrar especialistas ni conocimiento técnico, donde el administrador de contenido permite adaptarse a las tendencias en Internet

donde se está pasando de tener contenidos estáticos a contenidos dinámicos y sitios interactivos” (Ortegón, 2007). En este concepto se reconoce la figura del administrador de contenido, concepto revolucionario en la gestión de la información en las páginas dinámicas de la web. Dicho reconocimiento implica un llamado a los profesionales de la información para desarrollar un espacio necesitado de nuestras herramientas y habitualmente ocupado por especialistas de otros campos relacionados con la informática.

Para Tramullas, un CMS es **“un sistema informático (hardware y software) responsable de la colección, gestión y publicación de bloques (chunks) de información, llamados componentes de contenido”** (Tramullas, 2009).

A partir del análisis de estos conceptos, se puede advertir que con el transcurso de los años, en las definiciones se introducen aspectos en función de las ventajas de las aplicaciones en la praxis, especialmente en función de la comunidad usuaria.

Otro de los elementos a resaltar es la tendencia a la citación y adopción de conceptos generados en los primeros años de abordaje de los CMS (entre el 2000 y 2010). Razón por la que se especula que escaseen definiciones fechadas en los últimos cinco años.

De los conceptos manejados se anotan los siguientes puntos coincidentes:

- Un CMS es un sistema: significa que están compuestos por un conjunto de elementos que se relacionan e interactúan entre si, de modo tal, que le confieren entidad propia para formar un todo unificado. Por tanto, al afirmar que un CMS es un sistema informático³ (Hernández and Pérez,

³ Un software o sistema informático, programa de cómputo, etc., debe entenderse, no solo como una secuencia de instrucciones, sino también a la documentación auxiliar y **la descripción de la misma, acotando que ellas comprenden “una representación completa de las instrucciones, suficiente para determinar el conjunto de operaciones que constituyen el programa”** HERNÁNDEZ, G. & PÉREZ, O. 2004. ***Software propietario y software libre: una cuestión de equilibrio*** [Online]. Available: <http://www.bibliociencias.cu/gsd/collect/eventos/index/assoc/HASH3162.dir/doc.pdf> [Accessed 18 abril 2009].

2004), se asume que como sistemas, tienen entradas y salidas a partir de aplicaciones informáticas.

- Son aplicaciones, software integradores para la creación de productos para la Web.
- Se reconoce al CMS como sistema dinámico y su actividad está en función de la creación, gestión y actualización de otro producto.
- Su ventaja esencial como herramienta para la construcción de productos Web, es que permite manejar el diseño (desde una perspectiva visual y de requerimientos de funcionamiento informáticos), de manera independiente de la gestión de los componentes del contenido (arquitectura, creación, almacenamiento, modificación, edición).
- Su estructura tecnológica opera sobre páginas dinámicas, las que se caracterizan por ser cómodamente actualizables desde un panel de control, con la ventaja de que para ello no es preciso conocer lenguajes de programación Web.

Como otros elementos que los distinguen del resto de las herramientas informáticas que coexisten en la construcción de productos Web, puede añadirse que se basan en el almacenamiento del contenido en bases de datos y el desarrollo de módulos que permiten la gestión de dichas bases de datos e implementación integrada de todas sus funciones, desde un único software.

Por tanto, los CMS pueden definirse como ***herramientas informáticas, que funcionan sobre la base de una filosofía de holística, estructuradas sobre la base de módulos que permiten la creación, almacenamiento, actualización, recuperación y visualización de los contenidos, a distintos niveles de gestión y acceso, en función del producto implementado.*** La adecuada aplicación de un CMS permitirá proporcionar un alto grado de personalización para los usuarios de modo tal que sus principales comunidades consideran que el sitio ha sido diseñado expresamente para ellos.

Es un sistema sinérgico, dinámico e integral, que permite manejar de manera independiente el diseño visual, de la gestión del contenido.

Con relación a la acción humana, para los CMS es de suma importancia un **entorno de trabajo de colaboración** distribuido, ya que es donde se puede aprovechar todo su potencial.

Es válido aclarar que, como en todo concepto, el nivel de generalización es considerable. Dada la variedad de CMS existentes, no todos tienen el mismo nivel de amigabilidad. Esta característica dificulta el manejo eficiente de cada uno de ellos y el grado de especialización que se exige en cada caso. Es decir, **de manera particular cada uno se define a partir de sus requerimientos informáticos.**

1.2. Origen e historia de los CMS

La Sociedad de la Información se ha desarrollado en función de las necesidades y los avances tecnológicos. En este sentido, la gestión de contenidos se ha beneficiado de las técnicas y procesos de producción y edición de publicaciones digitales.

Los CMS tienen su origen en el campo de la Informática. No obstante, como herramienta para la construcción de distintos productos electrónicos, fundamentalmente en ambiente Web, su uso se ha extendido hacia los diferentes puntos de contacto que en la práctica tiene con otras ciencias.

Antes de la aparición, instauración y difusión de los CMS, la gestión de las web se llevaba a cabo con aplicaciones independientes: editores de texto HTML y de imágenes, gestores de bases de datos y algoritmos de programación a medida.

Existe consenso entre los estudiosos del tema al asociar el surgimiento de los CMS con la popularización de Internet durante la segunda mitad de los años 90 (Tramullas, 2005, Wilkoff et al., 2001, Cuerda and Minguillón, 2004, Pérez-Montoro Gutiérrez, 2005), en un momento en que proliferan portales de organizaciones (web e intranet) que necesitaban de continuas actualizaciones.

Entre las empresas pioneras en el campo de los CMS, se distinguen:

- Illustra Information Technology: En 1994, utilizaba como repositorio de contenidos una base de datos con la cual proveía a sus usuarios un entorno amigable para su creación intelectual.

- RedDot: En 1995 presentaron su CMS basado en una base de datos (Cuerda and Minguillón, 2004).

- CNET: En 1995, este sitio de noticias tecnológicas introdujo un sistema de administración de documentos y publicación y creó una compañía llamada Vignette, que abrió el mercado para los sistemas de administración de contenido comerciales.

Otro de los sistemas de gestión de contenido que tuvo su nacimiento y desarrollo en esta década fue TYPO 3, cuya aparición en el mercado ocurrió en el año 1997.

Sin embargo, aunque los primeros CMS surgen en la última década del siglo XX, el término entonces era desconocido. No es hasta el año 2000 que los CMS se popularizan a partir del uso de PHPNuke como herramienta para las comunidades de usuarios en Internet.

La popularidad de los CMS no solo se hizo notar en aplicaciones prácticas, sino también en el mundo editorial. Por ejemplo, según estudios realizados por la autora, las publicaciones en revistas de la corriente principal hospedadas en el Web of Science (WoS) ⁴, comienzan a aparecer a partir del año 2000.

⁴ TOPIC: (*CMS OR "content management system"*) Refined by: WEB OF SCIENCE CATEGORIES (INFORMATION SCIENCE LIBRARY SCIENCE)

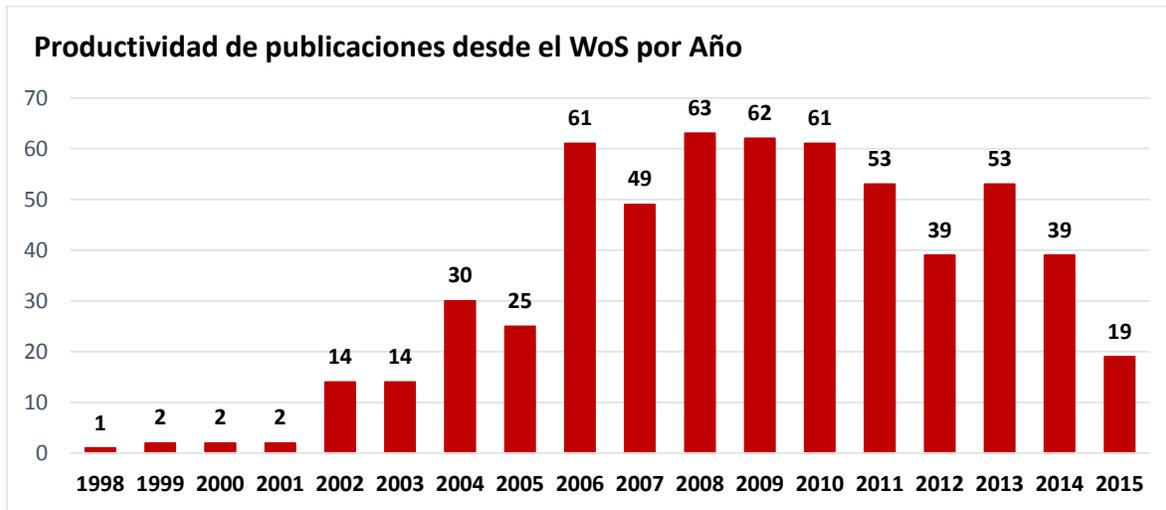


Fig. 1.2. Ejemplo del comportamiento de la frecuencia de publicaciones sobre CMS hasta el año 2015 (elaboración propia)

La aparición de los CMS en la Web fue un paso importante en la construcción de productos electrónicos, a mayor o menor escala. La evolución de Internet hacia portales con más contenido y la alta participación de los usuarios directamente, a través de blogs y redes sociales (web 2.0), los ha convertido en una herramienta esencial, tanto para empresas e instituciones como para las personas.

Como se menciona en párrafos anteriores, el proceso de creación, mantenimiento y/o desarrollo en la Web era una tarea que demandaba requisitos mínimos de especialización en programación, en correspondencia con el alcance del proyecto.

Bajo estas condiciones, la construcción, implementación y actualización de productos electrónicos para la Web, se volvía una tarea compleja para las empresas y las distintas instituciones de información. Esta idea fue compartida por Urra y Sarduy (2006), quienes distinguen como una de las causas a la que los CMS deben su aparición en el mundo de las Ciencias de la Información, la necesidad de suplir las insuficientes soluciones que brindaban los sistemas de información basados en páginas Web estáticas, pues requerían de especialistas dedicados a su desarrollo (Sarduy Domínguez and Urra González, 2006).

Los sistemas de información automatizados dependían esencialmente de un personal especializado y dedicado al procesamiento y gestión de la información a circular en el sistema. Como consecuencia, la información se demoraba en estar a disposición de los usuarios finales, al tener que circular por varios especialistas. Tal situación iba en contra de la efectividad del sistema porque estos procesos se tornaban demasiado complejos o trabajosos. Las organizaciones comenzaron a crear estrategias que estuvieran dirigidas a la integración de los recursos y a un flujo constante de los procesos de circulación de información, pero la solución a la necesidad que existía era más bien de corte tecnológico.

Estas ideas condujeron a la identificación de otra perspectiva en la aparición de los CMS. Santillán (2007), explica su surgimiento como herramienta a partir de la evolución y aplicación de los lenguajes de programación, en la creación de productos Web y las ventajas que representa el trabajo con páginas dinámicas. Para ello modela el siguiente esquema (Santillán, 2007):

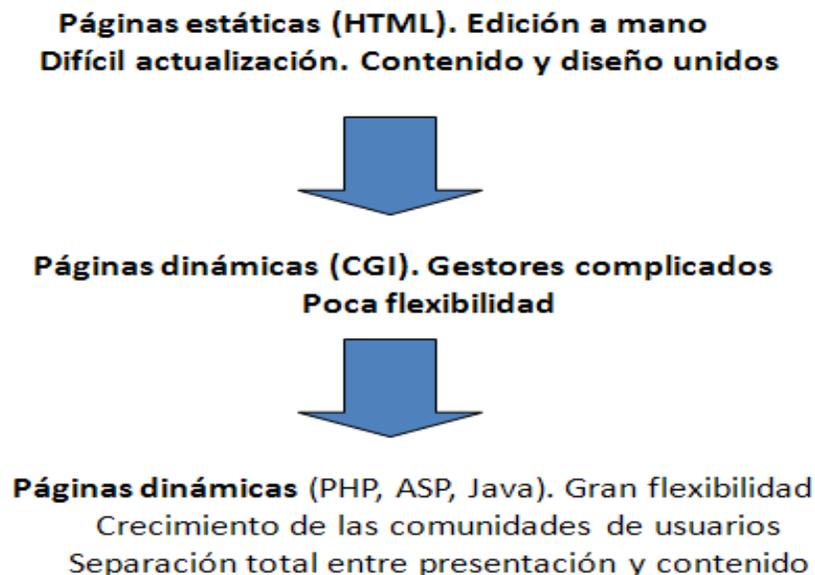


Fig. 1.3. Esquema sobre evolución de las páginas Web y los lenguajes que las sustentan (Santillán, 2007)

Las páginas estáticas se presentan sin movimiento y sin funcionalidades más allá de los enlaces, no interactúa con el visitante a través de alguna base de datos. Estas páginas son muy sencillas de crear, aunque ofrecen pocas ventajas tanto a los desarrolladores como a los visitantes, debido a que sólo se pueden presentar textos planos acompañados de imágenes y a lo sumo contenidos multimedia como pueden ser videos o sonidos. Una página Web estática presenta las siguientes características:

- Ausencia de movimiento y funcionalidades.
- Realizadas en XHTML o HTML.
- Facilitan el posicionamiento⁵ Web
- Extenso soporte en todo el mundo
- Portabilidad, funcionan en cualquier servidor
- Para cambiar los contenidos de la página, es imprescindible acceder al servidor donde está alojada la página.
- El usuario no tiene ninguna posibilidad de seleccionar, ordenar o modificar los contenidos o el diseño de la página a su gusto.
- El proceso de actualización es lento, tedioso y esencialmente manual.
- No se pueden utilizar funcionalidades tales como bases de datos, foros, etc.
- Mínimos requerimientos técnicos para su operación

Las páginas dinámicas, son aquellas con las que se puede interactuar. Incluye un grupo de funcionalidades y para ello es necesario utilizar otros lenguajes de programación, aparte de HTML, como PHP, ASP, etc. La información de una página Web dinámica puede ser almacenada dentro de una base de datos.

⁵ Según los distintos sitios visitados, en las páginas dinámicas el paso de parámetros a través de la URL permite al sitio interactuar con el usuario en función de diferentes variables. El problema es que algunos buscadores tienen dificultades para indexar estos sitios con URLs dinámicas. Además, los buscadores consideran que estas son poco estables (cambian mucho) y en numerosos casos llevan a contenidos duplicados, por lo que en busca del posicionamiento Web en buscadores, las páginas estáticas son más recomendadas.

Una página Web dinámica tiene las siguientes características:

- Gran número de posibilidades en su diseño y desarrollo.
- El visitante puede cambiar el diseño o presentación de la página a gusto.
- En su realización se utilizan diversos lenguajes y técnicas de programación.
- Ayuda a reducir considerablemente el tiempo de actualización y mantenimiento del contenido del sitio Web, pues es más sencillo, sin necesidad de entrar en el servidor y ofrecen autonomía en la administración de contenidos usando un CMS.
- Los autores del contenido no requieren conocimientos técnicos.
- Permite un gran número de funcionalidades tales como bases de datos, foros, chat, etc.
- Pueden realizarse íntegramente con software de libre distribución.
- Por lo general cuenta con el apoyo de una comunidad de desarrolladores.
- Cuenta con soluciones prediseñadas de libre disposición (opciones add-on)

Según Tramullas, con la aparición de Internet y los servicios que en ella se ofrecen, se ha ido pasando progresivamente de un concepto de publicación de páginas Web, bastante simple en su origen, a esquemas más complejos y diferenciados, fundamentados en procedimientos y técnicas basados en la gestión de información (Tramullas, 2005).

“La aparición de herramientas de gestión de contenidos vinieron a enmendar los problemas que presentaban los sistemas de información y llevaron a que las instituciones basaran sus sistemas de información en sistemas de gestión de contenidos”(Sarduy Domínguez and Urra González, 2006). Por tanto, entre las principales ventajas de los sistemas de gestión de contenidos está el permitir que, sin renunciar a los necesarios controles de calidad, cualquier persona, sin conocimientos avanzados sobre Informática, pueda colocar, modificar o eliminar contenidos del sitio.

Sin embargo, las crecientes demandas e intereses por mejorar los sistemas de búsqueda, recuperación y pertinencia de la información en la red, unido al

surgimiento del concepto de Web semántica, han hecho que los CMS, aunque **representan** “el paradigma actual en el diseño y desarrollo de sistemas y servicios de información basados en la Web” (Pastor Sánchez and Martínez Méndez, 2009), **necesiten implementarse sobre** “estructuras coherentes (...)”, con el fin de no poner trabas al acceso a la información y evitar la desubicación de los usuarios en sus procesos de navegación para la recuperación de información. Esto precisa del uso de herramientas próximas a la organización del conocimiento” (Pastor Sánchez and Martínez Méndez, 2009) como los tesauros, taxonomías, ontologías, etc.; reforzando la importancia de la organización de la información como elemento clave del trabajo con los CMS en la actualidad.

Por otra parte, ha cambiado la visión de los contenidos Web con tendencia a la definición de estructuras que establecen una tipología de contenidos con una semántica más explícita y cercana al concepto de metadatos (Horrocks, 2003, Horrocks et al., 2003) que permitiría un desarrollo más eficaz de la Web semántica.

Otra de las formas de mostrar la evolución de los CMS, fue presentada por Pastor, Orduña y Saorín. **Señalan que estas herramientas** “han alcanzado durante la última década un alto grado de desarrollo conceptual desde el punto de vista de la arquitectura de la información, la interconexión con bases de datos corporativas, la modularidad y la reutilización de contenidos. La disponibilidad en abierto de muchas de estas aplicaciones y su facilidad de instalación y manejo han propiciado un uso su facilidad de instalación y manejo han propiciado un uso”, **aunque todavía queda mucho por hacer en el campo** de la integración de los contenidos (Pastor Sánchez et al., 2013).

Según dichos autores la evolución de los CMS puede resumirse en tres etapas esenciales:

1ª fase: contenido y presentación mezclados;

2ª fase: contenido y presentación separados;

3ª fase: contenido, presentación y significado separados.

Tomando las diferentes visiones teóricas en consideración, podría ampliarse el esquema evolutivo propuesto por Santillán (2007), quedando la siguiente propuesta.

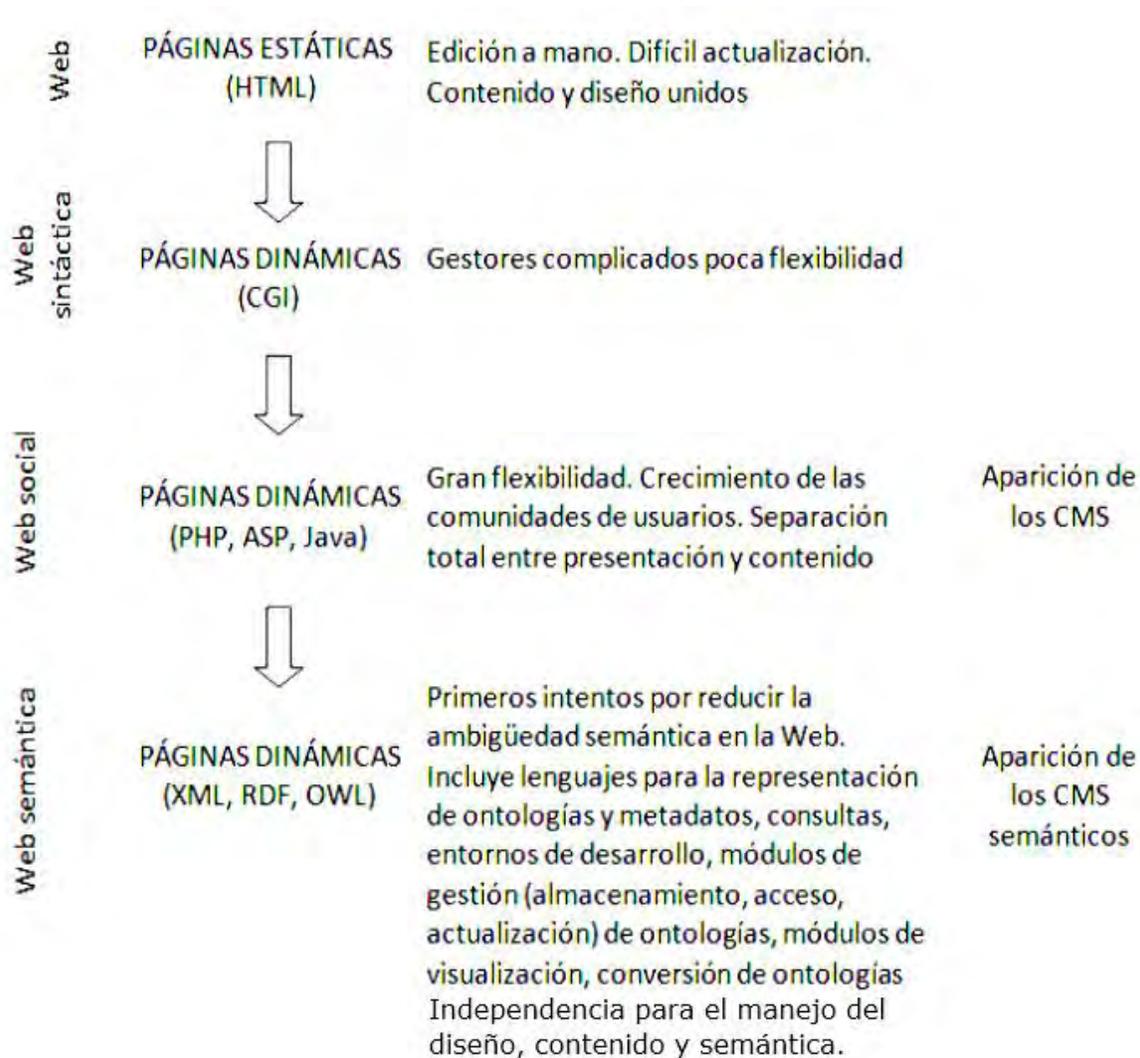


Fig. 1.4. Esquema evolutivo de los CMS y su relación con el desarrollo de lenguajes de programación para la Web (Elaboración propia, basado en el esquema de Santillán, 2007)

Se coincide con Robertson cuando, sobre el futuro de los CMS, apuntaba que este campo va en vías de alcanzar un alto grado de consistencia y profesionalismo que los proyectos dirigidos a la gestión no podrán ignorar (Robertson, 2003).

1.3. Estructura de los CMS: modelación conceptual y principales componentes

El estudio de la estructura de los CMS puede ser enfocada desde varias aristas. Desde una dimensión teórica, distintos autores han modelado el funcionamiento del CMS, basados funciones básicas, primarias, partiendo del principio de generalización que permite desde la teoría una mejor comprensión del objetivo de este tipo de software.

Desde la dimensión tecnológica, se torna más compleja la modelización. Bajo este enfoque, por lo general, se identifican parámetros que los definen como software.

Estructura desde la perspectiva de modelos conceptuales:

La estructura de los CMS modelada desde una perspectiva teórica, se basa en la identificación de estructuras, funciones y establece relación entre ellas.

Boiko (2001) señala que un CMS se compone de diferentes subsistemas que interactúan (Boiko, 2001):

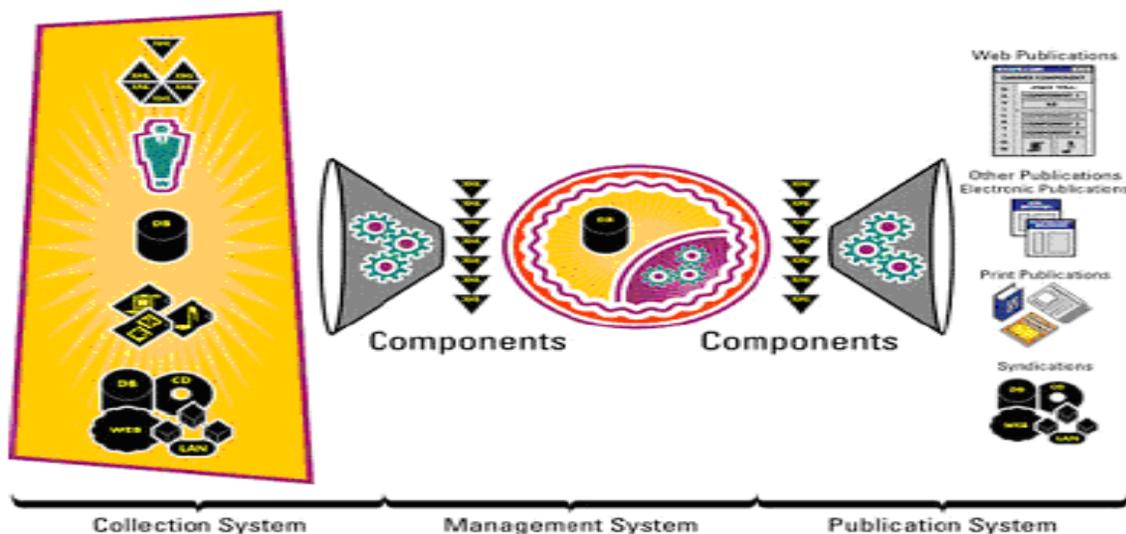


Fig. 1.5. Estructura del CMS según Boiko (Boiko, 2001)

- **Sistema de colección:** Responsable de todos los procesos que se dan antes de que el contenido sea publicado: creación, adquisición, conversión, agregación, adición de metadatos, etc. Es decir, soporta los procesos de

creación de contenidos y adquisición de la información, lo que permite dar soporte a los flujos de trabajo, sindicación e integración de fuentes externas.

- **Sistema de gestión:** Responsable del almacenamiento de los componentes que formarán los contenidos y otros recursos. Integrado por repositorio, el flujo de trabajo y las funcionalidades de administración. Por tanto, tributa al control y gestión de flujos de trabajo, grupos de usuarios y de la definición de parámetros para el funcionamiento del sistema.
- **Sistema de publicación:** Responsable de producir publicaciones de forma automática a partir de los componentes y otros recursos almacenados en el repositorio. Incluye, entre otras, las plantillas de visualización (publishing templates). Utiliza un modelo basado en templates, ofrece posibilidad de personalización de usuarios y diferentes tipos de plataformas.

McKeever (2003) plantea que, como modelo conceptual, un CMS sería una superposición de capas que interactúan con los contenidos (McKeever, 2003).

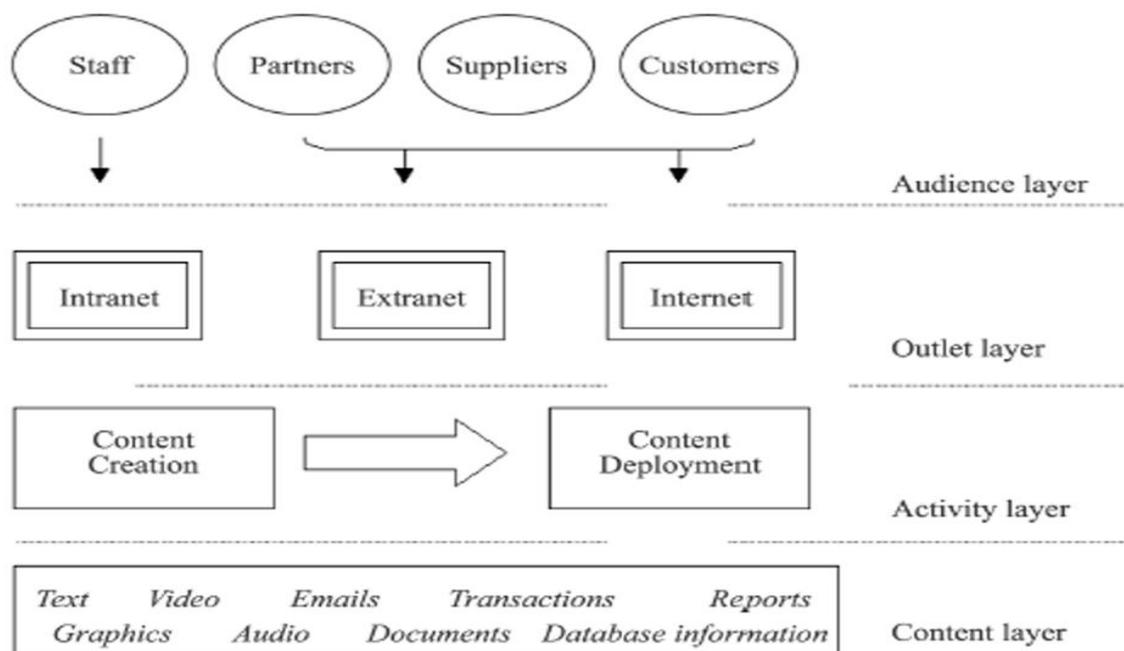


Fig. 1.6. Estructura del CMS según McKeever (McKeever, 2003)

En la primera capa encontramos los contenidos, se definen los tipos de contenidos. En la segunda, las actividades relacionadas con la creación y la implementación de los contenidos. En la tercera, las actividades relacionadas con la publicación, por ejemplo, los espacios de salida de los contenidos: web o intranet. En la cuarta, las actividades relacionadas con la definición de las audiencias: se establecen los usuarios posibles que interactuarán con los contenidos publicados. En la quinta, la definición de los dispositivos de salida: ordenadores, impresoras, etc.

Robertson (2003) modela la estructura del CMS a partir de cuatro funciones, que para el autor son la esencia de este software (Robertson, 2003).

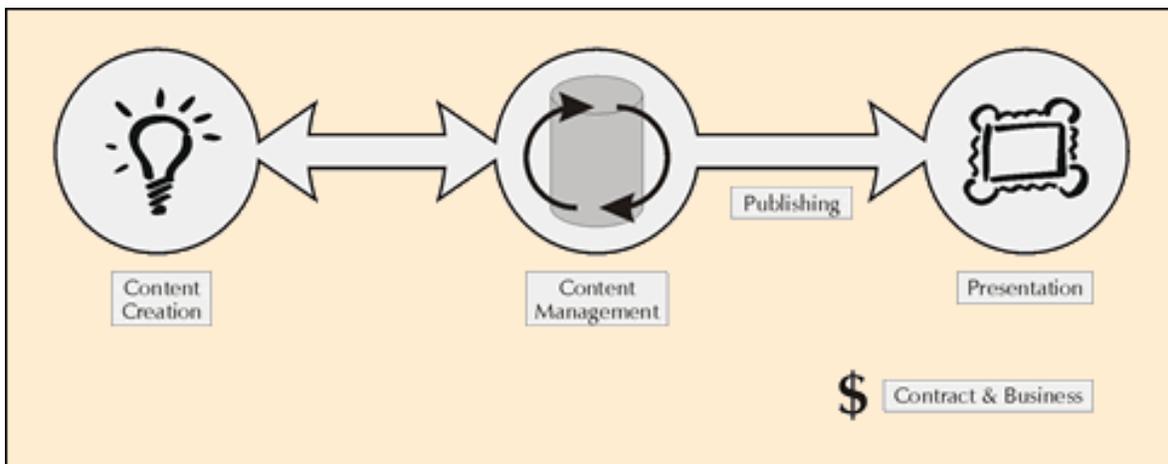


Fig. 1.7. Estructura del CMS según Robertson (Robertson, 2003)

- **Creación de contenidos:** Abarca las funcionalidades relacionadas con la creación por parte de los autores y las herramientas para gestionar la estructura del sitio y enlazar unas páginas con otras.
- **Gestión de contenidos:** Contiene el repositorio o base de datos donde se almacenan los contenidos. Reúne funcionalidades como el control de versiones, el control de acceso, la integración con otros recursos y los sistemas y los flujos de trabajo.
- **Publicación:** Funcionalidad que permite la exportación del contenido del repositorio en una o más webs o intranets en el mismo momento de la publicación (separación de contenido y presentación hasta que el contenido

no se publica). Asimismo, permiten la publicación de un mismo contenido en múltiples sitios con presentaciones y apariencia diferente.

- **Presentación:** Funcionalidades que mejoran la calidad y eficacia (dinamismo e interactividad) del sitio web. Ejemplo: construyen automáticamente la navegación del sitio web a partir de la estructura que haya establecido el autor en el repositorio.

Michelinakis (2004) apunta que, aunque las características de los CMS varían en cuanto a requerimientos y funcionalidades a partir de las necesidades concretas de las instituciones que lo generan o utilizan, existen elementos comunes que le permiten modelar su estructura general.

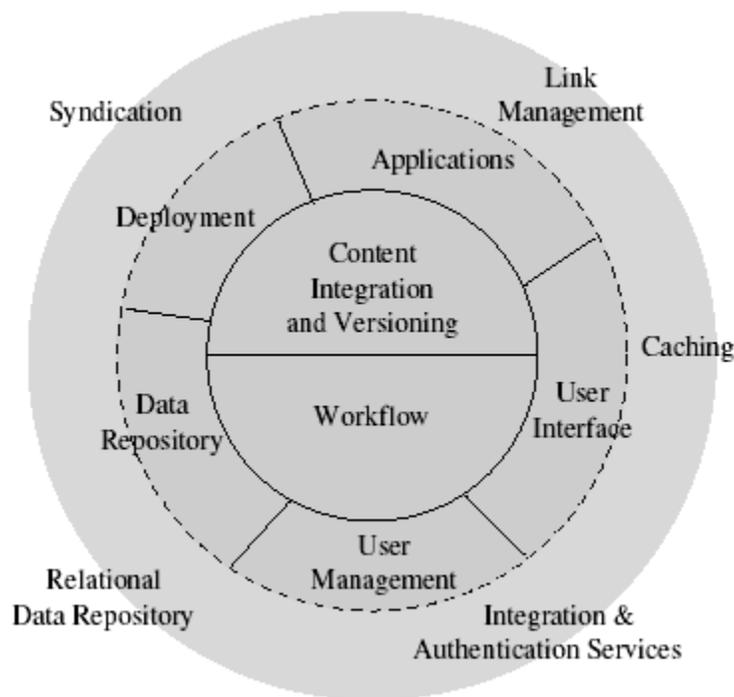


Fig. 1.8. Estructura del CMS según Michelinakis (Michelinakis, 2004)

La figura muestra una interpretación visual de la estructura típica de un CMS. La modelación tiene la llamada "característica de cebolla" (Michelinakis, 2004), por las capas que estos presentan. Dicha estructura es creada sobre la base de la estructura informática general del CMS. Cada una de las capas representa

distintos niveles. En el centro aparece el "núcleo", donde se definen las funcionalidades del CMS, y en la capa más externa el nivel de interacción con el usuario, los modos o formas de las entradas y salidas del sistema. La capa intermedia corresponde al almacenamiento del contenido. Pero el autor no ofrece explicación alguna sobre los criterios o modos de interacción de estos elementos.

Otros autores han sido mucho más exhaustivos en lo referido a la estructura de los CMS, por ejemplo, Jiménez y Moreo (2006). A partir de sus reflexiones (Jiménez and Moreo, 2006) se puede elaborar el siguiente esquema:

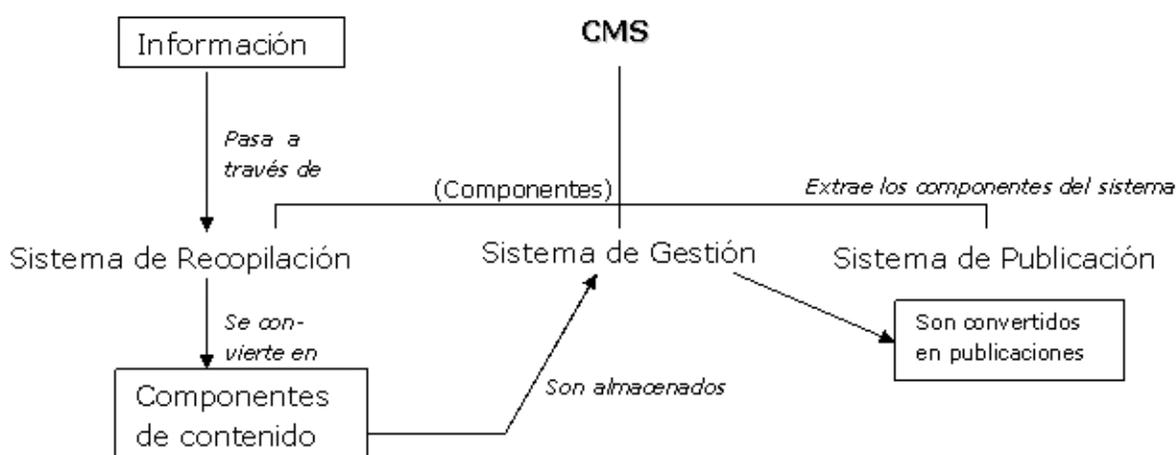


Fig. 1.9. Estructura de los CMS según la interpretación dada a las ideas de Jiménez y Moreo (elaboración propia)

Según lo plasmado, los CMS se componen de tres subsistemas esenciales (de Gestión, de Recopilación y de Publicación) que operan con los "componentes del contenido".

La información con la cual va a trabajar el sistema es captada y pasa al CMS a través del sistema de recolección y se convierte, una vez dentro, en componentes de contenido. Estos son almacenados más tarde en el sistema de gestión. Luego, a través del sistema correspondiente, los componentes de contenidos son publicados.

Por otra parte, Sánchez (2011) presentan la estructura genérica de un CMS a partir de tres capas: Cliente (o usuario), Servidor y Base de datos (Sánchez, 2011). (Ver Fig.1.10)

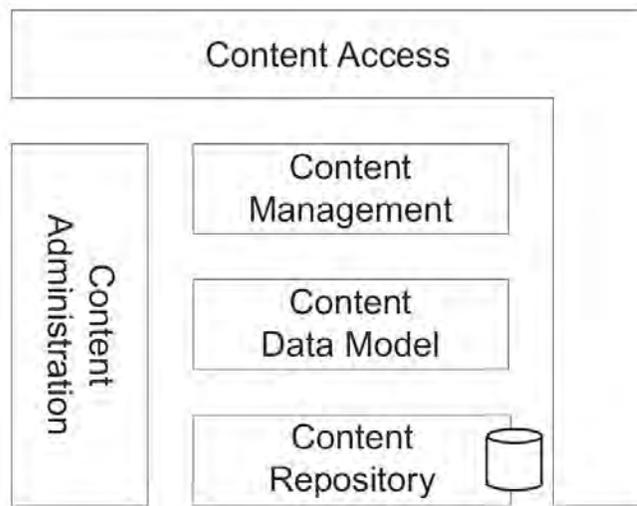


Fig. 1.10. Estructura de los CMS según (Sánchez, 2011)

A través de la interfaz de usuario del CMS (en la capa superior de la Figura 1.10: Acceso al contenido) se presenta el contenido y se ofrecen funciones de edición para crear, modificar y gestionar el contenido dentro de su ciclo de vida. Las funciones de administración central se implementan en la capa de gestión de contenidos. Esta capa proporciona funcionalidades para la definición del dominio o aplicación específica del *Content Data Model*. Control de acceso y las definiciones del ciclo de vida de contenido son otras características típicas de gestión implementados en esta capa. El contenido de la capa de modelo de datos se coloca conceptualmente debajo de la capa de gestión de contenidos que tiene las características necesarias para manipular el modelo. Esta capa de modelo de datos es el modelo específico de aplicación basado en el repositorio de contenido subyacente. El Repositorio de contenido define los conceptos fundamentales y los mecanismos de persistencia de datos para cualquier contenido modelo que se define en la parte superior. Las funciones de gestión de contenido están estrechamente relacionados con la capa de administración de contenido para administrar la pila CMS (Sánchez, 2011).

A partir de lo expuesto, se puede concluir que la diferencia entre los análisis hechos por los autores lleva a la identificación de dos tendencias en el modo de abordar el tema. Algunos, como Michelinakis, se refieren a la estructura con una visión informática del asunto, pero la mayoría de los autores consultados (Boiko, 2001, Robertson, 2003, McKeever, 2003, Jiménez and Moreo, 2006) se enfocan en las estructuras que permiten que ocurran los distintos procesos con la información. Todos coinciden al identificar estructuras, subsistemas o mecanismos que garantizan la entrada de la información al sistema (contenido) y la salida a través de la publicación y/o personalización de los contenidos. Los procesos del CMS como sistema son explicados, agrupados y analizados con diferentes niveles de exhaustividad, pero todos hacen referencia a la creación, gestión y visualización de los contenidos como funciones básicas de este software.

Estructura desde la perspectiva tecnológica:

Por otra parte, según un trabajo realizado por expertos (Ainoza Farled, 2007), existe una serie de parámetros que distinguen a los CMS y al mismo tiempo condicionan y ayudan a su selección para el desarrollo de sitios Web (CMS-Matrix, 2015)⁶. Los parámetros son: ***requerimientos del sistema, seguridad, soporte, usabilidad, gestión, interoperabilidad, flexibilidad, funcionamiento y aplicaciones.***

1. **Requerimientos del sistema:** Son las características generales que permiten hacer el proceso de selección inicial y entrada de los datos al sistema. Por tanto, para el proceso de selección de un CMS, es crucial definir el tipo de dato y los objetivos que se persiguen con el producto y que plantea la necesidad de una elección. Entre los elementos correspondientes a este

⁶ Aunque para el desarrollo del tema la autora se basó en el estudio antes mencionado, los elementos seleccionados, fueron comparados, analizados y respaldados por otras fuentes como el sitio Web CMS Matrix (<http://www.cmsmatrix.org>) y los que aparecen en el Web of Science, según estudio realizado por la autora como parte de su tesis de maestría, como autores que publican en el tema de los CMS.

parámetro se pueden enunciar las características del servidor en dos dimensiones: el **Servidor Web** y el **Servidor de Aplicaciones**.

El Servidor Web (ejemplos: Apache, IIS⁷) es la aplicación del software que suministra determinados archivos en respuesta a las peticiones de los navegadores Web, pero para ejecutar aplicaciones Web, debe funcionar con el servidor de aplicaciones.

A su vez, el Servidor de aplicaciones ayuda a un servidor a procesar páginas dinámicas. Este depende de factores como el presupuesto, el servidor Web elegido y su tecnología. Los servidores de aplicación pueden estar integrados en servidores Web o pueden descargarse gratuitamente de Internet (como, por ejemplo, Apache Tomcat).

Otro de los factores que distinguen a los CMS desde la perspectiva de los requerimientos son: el **Costo** aproximado del CMS (puede ser gratuito o no), la **Base de datos** que controla la gestión de contenidos del sitio (Por ejemplo: MySQL, Oracle, PostgreSQL), el tipo de **licencia** que permite el uso del sistema (por ejemplo: GNU GPL), el **Sistema operativo o plataforma** sobre el que se puede aplicar el CMS, el **Lenguaje de programación** (ejemplos: PHP, Python o PERL) y los **Privilegios** que permiten dar distintos niveles de interacción con el sistema, por ejemplo de administrador (**Root Access**).

2. Seguridad: Se refiere a aquellos elementos que, como el término lo indica, tributan a la seguridad, protección e integridad de los datos y el contenido que el CMS ofrece para proteger el producto creado.

Así pueden citarse: **Rastro de auditoría**, para dar la posibilidad de dejar un registro con todas las actividades dentro del sitio que desarrolla cada actor, la **Aprobación de contenidos** que asegura los mecanismos para el proceso de aceptación de los contenidos antes de mostrarlos en el sitio, la **Verificación por e-mail** para facilitar el registro del usuario con su dirección

⁷ Microsoft Internet Information Server

de correo electrónico, los privilegios “granulados” es decir niveles de libertad de acción según el usuario, *Historial de logins*, *Notificación de problemas*, *Gestión de sesión*, *Control de versiones* y distintos procesos de autenticación (por ejemplo: Autenticación de Keberos, Autenticación por LDAP⁸, Autenticación NIS⁹, Autenticación SMB¹⁰).

3. Soporte: Considerado un criterio para la selección por el CEIDS¹¹ en México. Es foco de atención tanto para los desarrolladores del CMS, como para los creadores del producto electrónico, pues juntos deben trabajar en la solución de los problemas. El soporte se refiere al CMS como software concretamente. Entre los elementos a contemplar se encuentra la existencia de *Programas de certificación* que permiten al usuario obtener los conocimientos y la técnica para usar el CMS, *manuales*, elementos de *formación* como cursos para formar al usuario, espacios donde se traten temas relacionados con la creación y desarrollo del CMS como una *comunidad de promotores*, espacios de *ayuda on-line* (foros, chats, etc.); Interfaz de programación de aplicación, *lista de correos y foros*, *Marco de pruebas* para el desarrollo y pruebas de calidad del sitio antes de que entre en funcionamiento.
4. Usabilidad: Se refiere a una serie de criterios y herramientas para la facilitar el uso del sistema o su aprendizaje. El CEIDS en México hace referencia al mismo criterio bajo el concepto de “Facilidad de Uso”. Se ven aspectos como la compatibilidad con Ms Internet Explorer y Firefox, *listas de discusión*, *URLs amigables*, *redimensión de imágenes*, *lenguaje macro* de programación para economizar tareas, posibilidad de hacer *transferencias masivas* de información (archivos, imágenes, sonidos, etc.), posibilidad de crear prototipos (*Prototipaje*) que sean modelos de la Web final, como una

⁸ LDAP: Lightweight Directory Access Protocol

⁹ NIS: Network Information Services

¹⁰ SMB: Server Message Block

¹¹ CEIDS: Centro Electrónico de Información e Investigación Documental para la Salud

forma de visualizar el sitio antes de crearlo lo que facilita la toma de decisión con respecto al diseño final del sitio.

Gingell incluye otros elementos como el verificador ortográfico, herramientas que se aplican a los procesadores de texto creador de estilos y lenguajes de plantillas (Gingell, 2003).

5. Gestión: Se refiere a las opciones de gestión y dirección que ofrece el CMS, para el sitio creado. Incluye elementos como: **gestión de archivos, de activos, portapapeles**, herramientas para la **planificación del contenido** y su organización según las preferencias del usuario, **Administración online o inline, iconos gráficos, papelera, estadísticas, gestión de estilos Web**, etc.

Existen otros parámetros que han sido tomados en consideración por otros estudios como:

6. Interoperabilidad: Capacidad del CMS para comunicarse con otros programas o sistemas.
7. Flexibilidad: Características que permiten adaptar el sistema a las necesidades del diseñador, desarrollador o usuario.
8. Funcionamiento: Herramientas que permiten el continuo desarrollo del CMS, para identificar puntos débiles y mejorarlos.
9. Aplicaciones: Se refieren a las funcionalidades de los CMS (gestión de documentos, gestión de FAQ, distribución de archivos, galería de imágenes, mapa del sitio, etc.) y el tipo de producto generado con él (ejemplos: blog, chat, foros de discusión, etc.)
10. Características relacionadas con el comercio, por lo general enfocadas a Internet como los pagos online, suscripciones, cesta de compra, punto de venta, etc.

CEIDS en México define otros criterios para la selección de un CMS, aunque comparte algunos de los ya mencionados.

11. Software de código (abierto o cerrado).

12. Arquitectura (robusta o simple).
13. Grado de desarrollo en el que se encuentra el CMS.
14. Posición en el mercado.
15. Accesibilidad.
16. Velocidad de descarga.

Todos estos elementos pueden usarse como patrones no solo para la caracterización de los CMS, sino también para su clasificación.

1.4. Tipologías de CMS

Dentro del campo de estudio de los CMS, como herramientas en constante desarrollo, hay distintas posibilidades y variantes de clasificación, ya sea por sus funcionalidades, casos de uso o por las tecnologías que se utilizan para crear las infraestructuras para la publicación y visualización de contenidos.

Entre los aspectos discutidos dentro del tema está la protección del software como una preocupación mundial. Su desarrollo requiere, como bien intelectual, un marco jurídico que le provea de cierta protección o tutela jurídica. En la actualidad, existen dos tendencias en cuanto a la propiedad del código del software:

- **Software propietario: Es asumido como "obra creadora, que debe estar sujeta a tutela judicial"** (Hernández and Pérez, 2004), pues se considera que la tutela legal es la que permite la defensa apropiada y con ello su estímulo y desarrollo. Han existido tres formas de protección: Derecho de patente, Sistema Sui Generis y el Derecho de autor.
- **Software libre: Se considera que el software debe estar libre y no sujeto a normativas. "Los partidarios de esta se vinculan a los principios de la Revolución Francesa: Libertad, Igualdad, Fraternidad, en la tecnología (...)** el software debe ser de libre acceso para todos y no debe estar sometido a ningún tipo de normativa que le reconozca derechos exclusivos y **capitalizadores a sus creadores"** (Hernández and Pérez, 2004). Se

distingue porque el código de fuente del programa está totalmente a disposición del usuario, es decir, una vez adquirido puede ser usado, copiado, estudiado, modificado y redistribuido libremente, sin ánimos de lucro.

En el ámbito internacional se mantiene el debate sobre el reconocimiento o aceptación del tipo de software libre, pero son muchos los que lo apoyan por las ventajas que ofrece sobre el software propietario en cuanto a costo y posibilidades de ajuste a las necesidades del usuario.

Un software libre debe respetar cuatro derechos o libertades considerados como fundamentales para los usuarios de ese programa (Stallman, 2004):

- Los usuarios deben tener derecho al programa, sin restricciones, donde quiera, como quiera y para lo que quiera.
- Los usuarios deben tener derecho a estudiar cómo funciona el programa y, si lo desean, a adaptarlo a sus necesidades específicas.
- Los usuarios deben tener derecho a distribuir copias a sus amigos, empleados, conocidos, empleadores y, en fin, a cualquier persona que deseen.
- Los usuarios deben tener derecho a mejorar el programa, publicar y distribuir sus mejoras al público (o a quien deseen) de modo que más personas salgan beneficiadas de los cambios. Por tanto, debe ofrecerse el código de fuente de la aplicación.

Como consecuencia de lo antes expuesto, la primera clasificación de CMS se podría dar desde una dimensión jurídica entre CMS propietarios y CMS de tipo software libre.

Cuando surgieron los sistemas de gestión de contenidos había que pagar por obtener las herramientas. Poco a poco se abrieron paso las de tipo software libre. Por lo general, los CMS propietarios son catalogados como CMS comerciales, pues son vendidos por empresas que consideran el código fuente un activo más que tienen que mantener en propiedad y no permiten que terceros tengan acceso.

De lo visto hasta el momento se puede modelar la dimensión jurídica del estudio de los CMS:

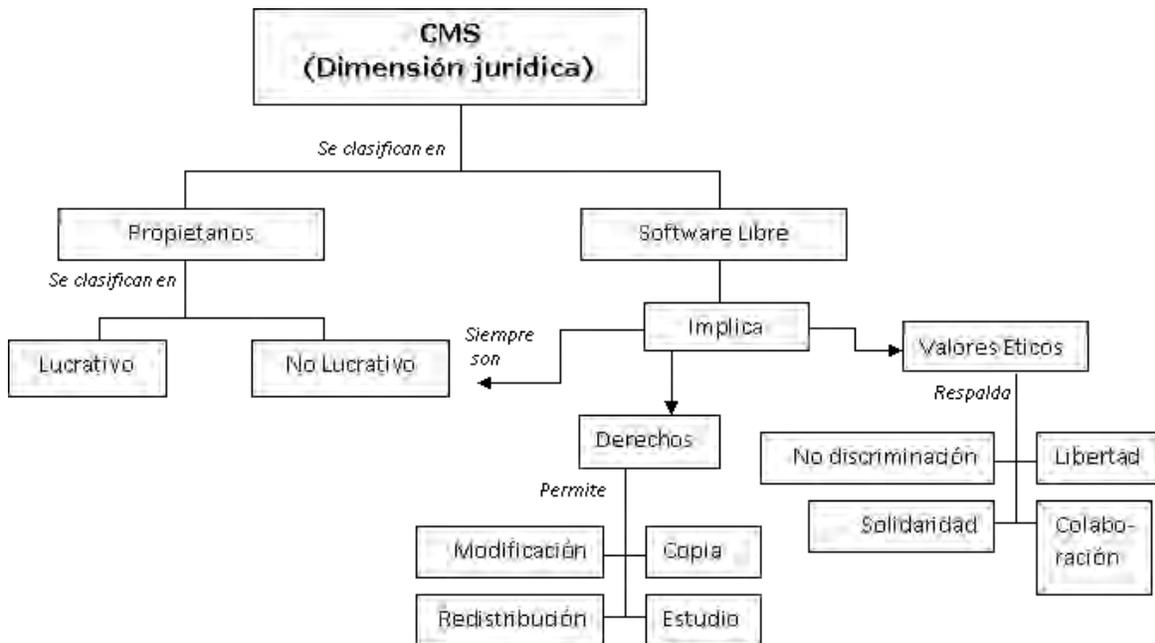


Fig. 1.11. Clasificación de los CMS desde la dimensión jurídica (elaboración propia)

Otra de las formas de analizar la clasificación de los CMS es a partir del contexto de uso, a partir de **“las diferencias existentes entre la gestión de contenidos para Web y la gestión de contenidos para empresas”** (Begoña Rodríguez et al., 2008).

Según Tramullas, la complejidad de estas herramientas ha generado dos especializaciones para la gestión de contenidos: Web Content Management (WCM) y Enterprise Content Management (ECM) (Tramullas, 2005). La primera, a decir de Miller y David (Miller and David, 2002), se centra en el entorno Web y su objetivo y métodos principales se orientan a la producción de documentos e información digitales para la Web, especialmente para portales. La segunda, según Gingell, se fundamenta en la gestión total de la información en las organizaciones (Gingell, 2003). **“Para ello integra toda la información necesaria para alcanzar los objetivos de la organización proveniente no sólo de sistemas de publicación, sino también de sistemas Enterprise Resource Planning (ERP), gestión de documentos, datawarehouses**

(...). **La expresión más generalizada actualmente del ECM son las intranets y los portales internos de las organizaciones**" (Osuna Alarcón and De la Cruz Gómez, 2010).

Desde otro ángulo de clasificación de los CMS, pueden agruparse según la infraestructura para la publicación, es decir, sus funcionalidades. Esta perspectiva de estudio, se desarrolla bajo una dimensión esencialmente tecnológica, pues son los requerimientos del sistema los que condicionan el tipo de producto que se genera.

Tramullas plantea que en los CMS se pueden encontrar herramientas y plataformas que muestran diferentes orientaciones, diferentes prestaciones y enfocadas a diferentes objetivos y grupos de usuarios. Propone su clasificación en: Plataformas para desarrollo de gestión de contenidos, Portales, Aulas virtuales, Bibliotecas digitales, Publicaciones digitales, Entornos para colaboración y Blogs o bitácoras (Tramullas, 2005).

A decir Díaz Domínguez los CMS pueden clasificarse atendiendo al tipo de sitio que se quiera gestionar: genéricos y específicos (Díaz Domínguez, 2010):

- Genéricos: Ofrecen la plataforma necesaria para desarrollar e implementar aplicaciones que pueden servir para construir soluciones de gestión de contenidos, de comercio electrónico, blogs, portales, etc.
- Específicos: Ofrecen una plataforma de servicios de Internet que en ocasiones incluye además del CMS herramientas para la recaudación de fondos para organizaciones sin fines de lucro, los interesados en la gestión de proyectos, etc. Entre ellos foros, Blogs, Wikis, e-Commerce, Portales, Galerías, E-Learning.

Otro de los estudiosos del tema, Boldú señala que los CMS se han especializado para dar respuesta a finalidades concretas, lo que ha dado lugar a la aparición de diferentes tipos de CMS en relación al tipo de contenido, el tipo de publicación y flujos de trabajo (Boldú, 2011). A partir de una sistematización teórica desarrollada por dicho autor, identifica cinco tipologías

esenciales, sin dejar de reconocer que no siempre la distinción es clara y que un CMS puede estar en diferentes categorías:

- CMS Genéricos: Orientados a la creación y publicación de webs e intranets con finalidades informativas y de servicios. Pueden integrar módulos para realizar funciones específicas, como galerías. El repositorio está preparado para gestionar todo tipo de contenidos. El flujo de trabajo es jerárquico. Ejemplos: Joomla, TYPO3, Drupal, Plone, Mambo.
- Blogs: Orientado a la creación y publicación de páginas personales, aunque algunas empresas lo utilizan para la difusión de información corporativa. No hay flujo de trabajo. El contenido se presenta de manera cronológica y por categorías. Pensados para que los usuarios aporten comentarios y generar, así, discusión y diálogo. El repositorio permite almacenar básicamente HTML, imágenes y vídeos. Han supuesto la democratización de Internet por su sencillez y simplicidad de funcionamiento y, sobre todo, porque se ofrecen como servicio totalmente gratuito para el autor: no necesita instalar ningún programa, ni pagar ningún servidor, ni dar de alta ningún dominio propio. Las entradas suelen ser comentarios personales, opiniones. Se explota la sindicación de contenidos (RSS, ATOM, etc.). Ejemplo: WordPress.
- Image CMS: Sistemas especializados en la gestión y publicación de imágenes (fotografía, arte, etc.). Permiten una categorización muy detallada de las imágenes mediante el uso de metadatos. Generan galerías automáticas en función de los metadatos, de la distribución de las imágenes en los directorios del repositorio y de las búsquedas de los usuarios. Funciones avanzadas de importación (empleando generalmente AJAX -Asynchronous JavaScript And XML-). Posibilidad de que los usuarios valoren las imágenes y descarguen originales de alta resolución. Generación automática de landingpages: últimas imágenes añadidas, las más votadas, las más vistas, las vistas recientemente, las comentadas recientemente, etc. Posibilidad de visualizar galerías y álbumes tipo presentación (slideshows). Posibilidad de edición y

tratamiento en línea de las imágenes importadas (recortar, ampliar, efectos y filtros, etc.) Creación automática de marcas de agua.

- e-Commerce CMS: Sistemas especializados en la gestión de tiendas virtuales o comercio electrónico. Los contenidos son productos y servicios. Catálogo online y galerías de productos. Tienda virtual con carro de compra, pagos por Internet (PAYPAL, etc.), Facturación, y envío de pedidos. Control de stocks.
- e-Learning CMS: "Sistemas especializados en la creación y gestión de "campus virtuales". Los contenidos son "learning objects". Administración de usuarios avanzada, que permite la creación de cursos / asignaturas + profesores y estudiantes / participantes. Espacio propio de creación de contenidos con funcionalidades de un CMS genérico. Herramientas de creación de actividades docentes (ejercicios, etc.) Entrega y calificación de trabajos, exámenes y tests con valoraciones, etc. Registro de la actividad de los usuarios participantes. Aparición de estándares propios: IMS Global Learning, SCORM, LOM, etc. Ejemplo: Moodle, Claroline.

Otra de las dimensiones para la clasificación de los CMS viene dada por los procesos de Organización¹² y Recuperación¹³ de la Información (ORI). En los últimos años, ha crecido la popularización del paradigma de la web semántica, con un marcado interés por resolver las limitaciones en los mecanismos de procesamiento automático para manejar grandes cantidades de contenido y **páginas, así como la interoperabilidad entre ellas.** "La semántica documental ha conformado tradicionalmente un pilar esencial en la organización y recuperación de la información. La irrupción de la Web y su evolución hacia una Web Semántica y de datos enlazados exige la adaptación de estas herramientas y vocabularios documentales hacia nuevos estándares de

¹² En el presente informe se asume que el término de organización incluye los procesos de procesamiento (catalogación, indización y resumen), representación y ordenación de los contenidos.

¹³ En el presente informe se asume que el término recuperación incluye los procesos de búsqueda y recuperación de los contenidos.

representación” (Alvite Díez, 2013). En dicho contexto se avizora un auge, dentro del gremio de los profesionales de la información, en el uso de la clasificación de CMS semántico para dar crédito a la característica de los CMS con posibilidades de un mejor entendimiento del lenguaje por parte de las máquinas. De esta manera y por inferencia, puede derivarse otra de las formas de agrupación de los CMS, tal como se presenta en la figura 12:

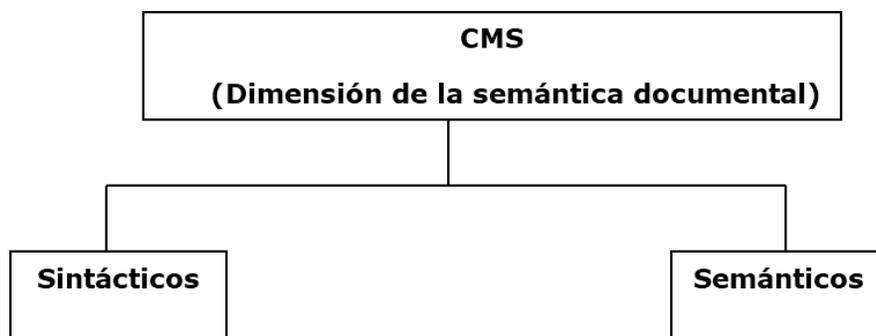


Fig. 1.12. Clasificación de los CMS desde la dimensión de la semántica documental (elaboración propia)

Entre los CMS sintácticos se encuentran, por ejemplo Plone, Apache Lenya, Mambo, PHP-Nuke entre otros. Son aquellos que, aunque permiten la sindicación de los contenidos a partir de los RSS para la indización automatizada, no reconocen la polisemia, la sinonimia ni las relaciones significativas que dan sentido semántico al contenido para las opciones e inferencias en recuperación de la información. Basan sus esquemas de representación en la recuperación de información a través de las relaciones entre documentos y/o sus partes, no en las deducciones o derivaciones relacionales de los contenidos, es decir, las implicaciones semánticas.

Por otra parte, los CMS semánticos¹⁴ son aquellos que se apoyan en tecnologías semánticas para la descripción de los contenidos, como por ejemplo RDF, OWL y XML. Estas tecnologías se combinan para aportar descripciones explícitas de los recursos de manera que sean entendibles por un software (relacionando e infiriendo datos y cosas, no solo páginas). Esas

¹⁴ Se abordarán con mayor profundidad en el Capítulo 3

etiquetas permiten que los gestores de contenidos interpreten los documentos y realicen procesos inteligentes de captura y tratamiento de información. Dentro de este grupo puede citarse como ejemplo a JOOMLA, DRUPAL, WordPress entre otros.

A partir del análisis de las principales ideas de los autores estudiados, se advierte que la clasificación de los CMS se realiza desde cuatro dimensiones: jurídica, social, tecnológica y de la semántica documental. Así se modela en la figura que se muestra a continuación.

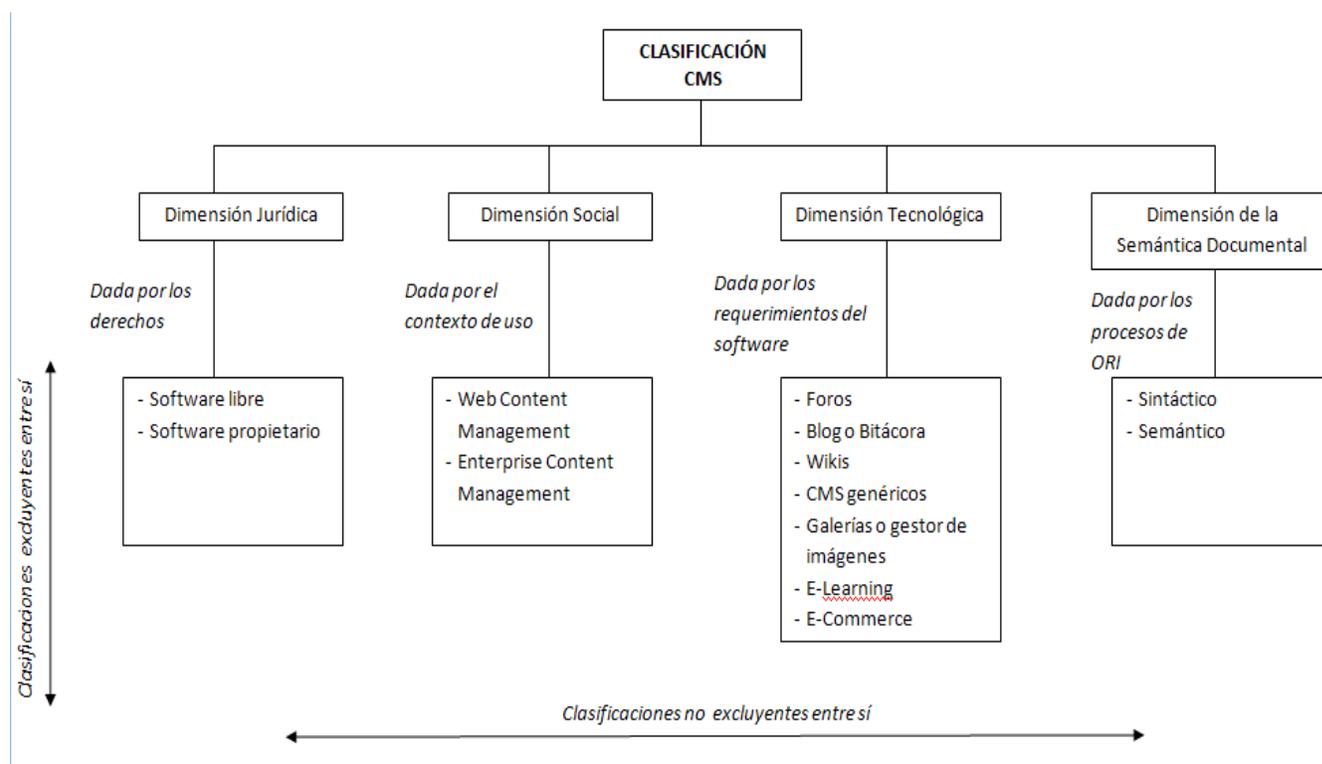


Fig. 1.13. Esquema resumen sobre la clasificación de los CMS (elaboración propia)

Desde cada una de ellas se establecen diferentes agrupaciones. Excluyentes entre sí hacia el interior de la misma y en el sentido horizontal del esquema son categorías perfectamente superpuestas.

1.5. ¿Por qué estudiar los CMS desde las CI?

La CI, distinguida por su carácter integrador, se ha nutrido de las aportaciones que, desde otras disciplinas, complementan el estudio de la información como fenómeno social. Al buscar y ampliar el conocimiento, para dar respuesta a los

retos como especialidad, establece nexos con otros campos con el objetivo de lograr un acercamiento al mismo fenómeno desde distintos ángulos.

Según Pérez y Setién, uno de los primeros acercamientos hacia la condición interdisciplinar de la Ciencia de la Información lo constituye un artículo de Taylor (1966), a pesar de que en la época en que se desarrolló su discurso, esta ciencia se encontraba en su fase más joven de estudio y formación (Pérez Matos and Setién Quesada, 2008).

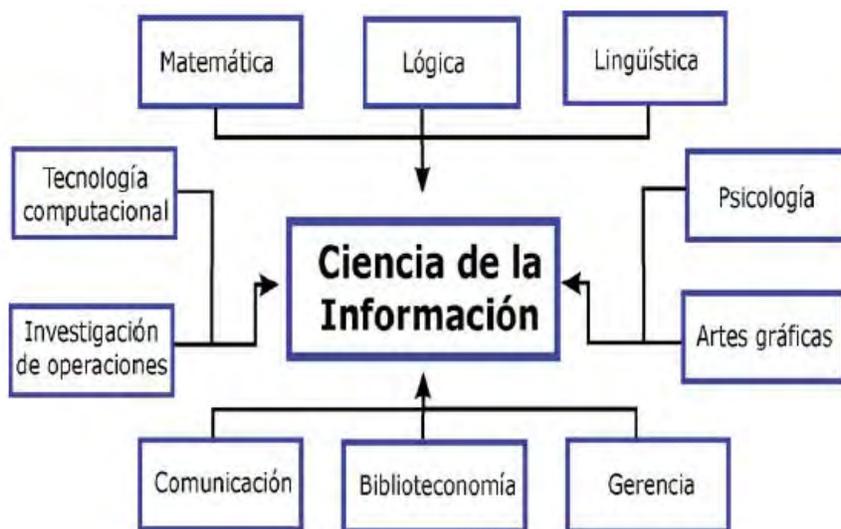


Fig. 1.14. Interdisciplinariedad sistémica integrada de la Ciencia de la Información (Taylor, 1966)

El esquema evidencia la concientización de la necesidad de la integración con otras disciplinas para el complejo estudio de los procesos informacionales (Taylor, 1966). Desde entonces se avizoraba la estrecha relación con disciplinas como la Lógica, la Lingüística y las NTIC, espacios que tributan directamente al desarrollo de las bases teóricas de los estudios de los CMS, desde la perspectiva de la CI.

Con el desarrollo de la CI como ciencia, el grado de complejidad y profundización con relación a su carácter inter, multi y transdisciplinar, aumentaron considerablemente.

La idea de la dimensión disciplinar de la CI fue defendida por Pinheiro quien define cinco categorías de disciplinas para esta ciencia: Estructurales,

Instrumentales o de representación, Gerenciales, Tecnológicas y Socioculturales (Pinheiro Ribeiro, 2006). De igual forma modela un esquema donde se representan diferentes sub-áreas y su interacción con otras disciplinas.

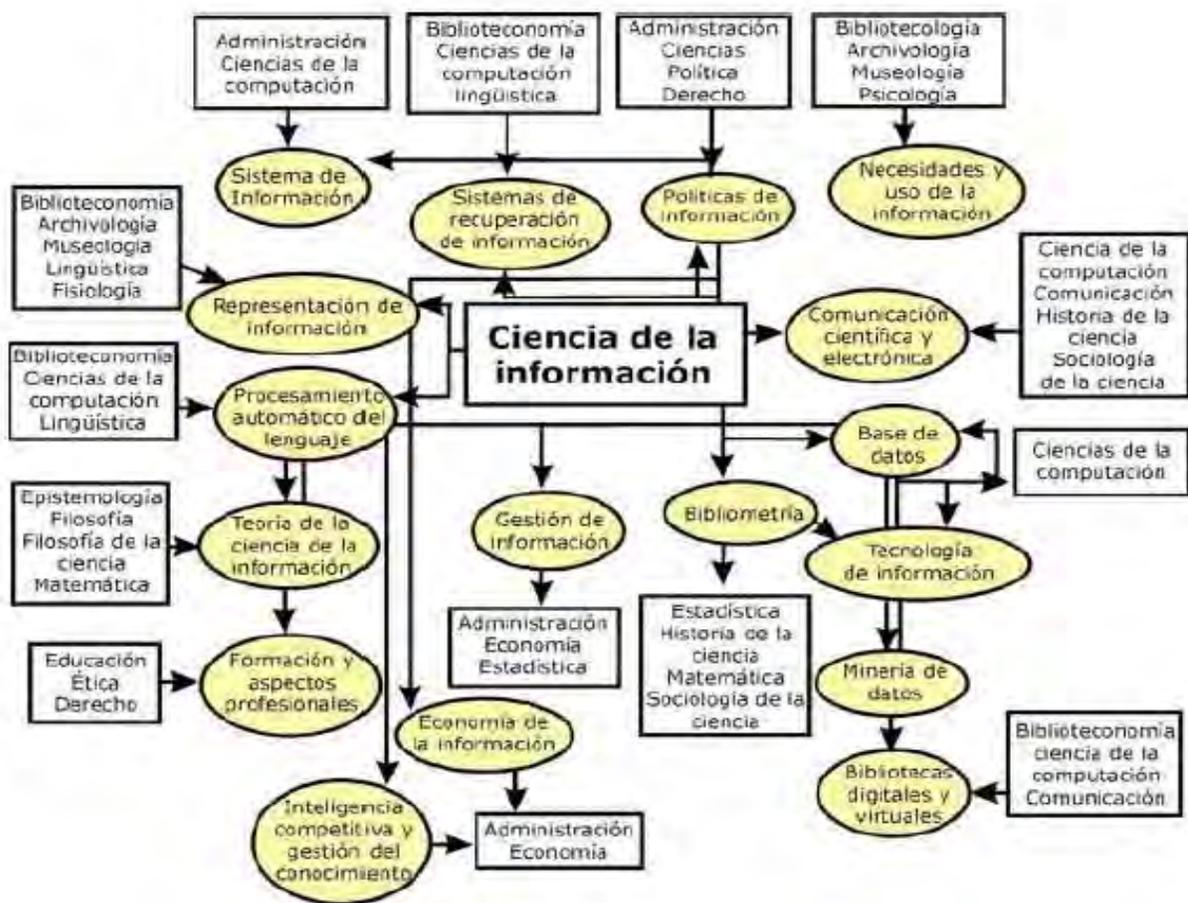


Fig. 1.15. Interdisciplinariedad lineal de la Ciencia de la Información (Pinheiro Ribeiro, 2006)

El avance de la CI, dedicada al estudio de los fenómenos informacionales, implica la yuxtaposición de dominios cognitivos muy diversos. Como principio del esquema, los fenómenos informativos mantienen diferentes niveles de complejidad e imbricación. El papel de las NTIC, en cada uno de estos espacios, es cada vez mayor, como consecuencia lógica y directa del desarrollo científico técnico.

El fuerte componente tecnológico, social y económico de la Sociedad de la Información, ha dado lugar a la incorporación progresiva de sus profesionales

en los nuevos ámbitos de tratamiento, gestión y difusión de información, especialmente digital. Dentro de este campo, los CMS ofrecen la posibilidad de gestionar de manera eficaz los contenidos digitales, así como facilidades para la creación e implementación de nuevos servicios. Por tanto, el profesional de la información no puede dar la espalda a una realidad que se impone dentro de su radio de acción. Por ende, es convocado, una vez más, a asumir un papel de mediador entre las tecnologías de la información y el usuario.

Varios autores han hecho alusión a los cambios socio-económicos en el desarrollo de la sociedad que demandan la evolución en los modos de hacer y proyectar los servicios desde la CI. En tal sentido Portela **plantea que "hemos pasado de la economía industrial a la economía del conocimiento, de la sociedad de la información a la informacional. Internet es una realidad y su potencial es enorme. A los profesionales de la gestión de información, Internet nos ha abierto nuevas áreas de trabajo, de promoción, de adaptación a la nueva realidad social y económica: creación y diseño de Webs, portales, intranets, intermediación y venta de contenidos son algunas de esas nuevas áreas"** (Portela, 2001).

Bajo la misma lógica de pensamiento Osuna y Cruz (2010), identifican distintos contextos ocupacionales del profesional de la información en el Web: editor web, especialista en búsquedas, redactor de contenidos, adaptador web¹⁵ y testadores¹⁶. Según opinión de estas autoras, del profesional de la información en el Web, en un entorno o sistema de gestión de contenidos, podrían resumirse en cuatro tareas, todas ellas propias del proceso documental: incorporación de contenidos, edición y publicación, recuperación y evaluación y mantenimiento de los contenidos.

¹⁵ **"Es el encargado de corregir y adaptar los contenidos de un servidor, un portal o una página web (...) Su trabajo es la edición, corrección, documentación y demás aspectos relacionados con los contenidos propios de la páginas web."**(OSUNA ALARCÓN, M. R. & DE LA CRUZ GÓMEZ, E. 2010. Los sistemas de gestión de contenidos en Información y Documentación. *Revista General de Información y Documentación*, 20, 67-100.).

¹⁶ **"Su trabajo se basa en la navegación por todo el sitio web para detectar cualquier posible fallo antes de su publicación"** (ibid.).

Dichas tareas coinciden con las principales funcionalidades de los CMS. Por tanto, desde la CI se impone la necesidad de estudiar cómo se desarrollan estos procesos, para calzar las debilidades que desde el punto de vista informacional puedan darse en el ciclo de vida de la información del producto Web que se genera. Pueden citarse, además, otras áreas de acción que competen al profesional de la información: arquitectura de información, estudios de usabilidad¹⁷, socialización y difusión de la información a partir de estudios de necesidades, trabajo en función de garantizar la fiabilidad y veracidad e integridad de la información, inculcar los principios del derecho de autor, alfabetización informacional, gestión documental, gestión de información, etc.

Si el estudio y trabajo con los S-CMS se deja únicamente en manos de profesionales provenientes del campo de la informática, en los productos que hoy se construyen no se encontraría respuesta a los procesos que atañen a la calidad de la información gestionada.

Los profesionales de la información deben asumir un papel activo, pues si los CMS facilitan el manejo de funcionalidades y acciones de programación como una de sus principales virtudes, entonces la profesión está llamada al estudio y enriquecimiento de los procesos que responden al tratamiento semántico del contenido que, a su vez, condiciona y caracteriza los flujos de trabajo generados en este contexto (procesamiento, representación, evaluación, estudio de necesidades de usuario, conservación de la información, etc.) lo que llevaría a elevar eficacia y eficiencia de los productos Web implementados en los diferentes espacios.

Por tanto, la CI desde sus distintas disciplinas, puede contribuir al estudio, perfeccionamiento y aprovechamiento óptimo de estas herramientas tecnológicas. El proceso de construcción de su marco teórico, al asumirse

¹⁷ Los estudios de usabilidad permiten evaluar la efectividad de un sitio web a partir de criterios basados en las facilidades que proporciona el sitio para alcanzar los objetivos que persiguen sus usuarios al visitarlo

desde esta ciencia, debe ser analizado desde la complejidad, precisamente por la multidisciplinariedad del tema.

1.6. Referencias bibliográficas

- AINOZA FARLED, J. 2007. *Estudio y evaluación de sistemas gestores de contenidos web*. Tesis de maestría, Universitat Politècnica de Catalunya. [Online]. Available: <http://hdl.handle.net/2099.1/5276> [Accessed 18 de enero de 2016].
- ALVITE DÍEZ, M. L. 2012. El uso de vocabularios controlados en los sistemas de información jurídica: evolución y tendencias actuales de representación. *Scire*. 18 (1), 29- 39.
- BEGOÑA RODRÍGUEZ, M., BAZÁN, P. & HUNGARO, A. M. 2008. *Espacios virtuales con valor educativo*. *Espacios virtuales con valor educativo* [Online]. Available: http://www.linti.unlp.edu.ar/uploads/docs/espacios_virtuales_con_valor_educativo.pdf [Accessed 24 de marzo de 2013].
- BOIKO, B. 2001. Understanding content management. *Bulletin of the American Society for Information Science and Technology*, 28, 8-13.
- BOLDÚ 2011. Curso: "Sistemas de gestión de contenidos Digitales (SGCD): Tema 4. TIPUS DE CMS". En: Màster en Gestió de Continguts Digitals (curso 2010-2011). Barcelona, España: [notas de clase].
- BRAMSCHER, P. F. & BUTLER, J. T. 2006. LibData to LibCMS - One library's evolutionary pathway to a content management system. *Library Hi Tech*, 24, 14-28.
- CMS-MATRIX. 2015. *Content Management Directory Matrix* [Online]. Available: <http://www.cmsmatrix.org/> [Accessed 24 diciembre 2015].
- CUERDA, X. & MINGUILLÓN, J. 2004. Introducción a los Sistemas de Gestión de Contenidos (CMS) de código abierto. *Mosaic*, 36. [Online]. Available: <http://www.uoc.edu/mosaic/articulos/cms1204.html> [Accessed 24 de mayo de 2010].
- DÍAZ DOMÍNGUEZ, N. P. 2010. *Los sistemas gestores de contenido y sus aplicaciones*. Tesis de licenciatura, Universidad Veracruzana.
- GINGELL, D. 2003. *A 15 Minutes Guide to Enterprise Content Management* [Online]. Available: http://www.aiim.org.uk/publications/ecm_at_work/pdfs/Ecm_15min_guide.pdf [Accessed 24 abril 2013].
- HERNÁNDEZ, G. & PÉREZ, O. 2004. *Software propietario y software libre: una cuestión de equilibrio* [Online]. Available: <http://www.bibliociencias.cu/gsd/collect/eventos/index/assoc/HASH3162.dir/doc.pdf> [Accessed 18 abril 2009].

- HORROCKS, I. 2003. Rules for the Semantic Web. *Ieee Intelligent Systems*, 18, 76-77.
- HORROCKS, I., PATEL-SCHNEIDER, P. F. & VAN HARMELEN, F. 2003. From SHIQ and RDF to OWL: The making of a web ontology language. *Web semantics: science, services and agents on the World Wide Web*, 1, 7-26.
- JIMÉNEZ, M. & MOREO, M. 2006. *Diseño de portales mediante CMS*. [Online]. Available: <http://lsi.ugr.es/~jparets/Escritos/Dise%F1o%20de%20portales%20mediante%20CMS.pdf> [Accessed 1 de abril de 2009]
- MCKEEVER, S. 2003. Understanding web content management systems: evolution, lifecycle and market. *Industrial Management & Data Systems*, 103, 686-692.
- MICHELINAKIS, D. 2004. *Open Source Content Management Systems: An Argumentative Approach*. Tesis de maestría, University of Warwick.
- MILLER, B. & DAVID, D. 2002. *Directions in Web Content Management*. [Online]. Available: <http://www.ddavid.com/groundzero/Directions%20in%20Web%20Content%20Management.pdf> [Accessed 23 de abril de 2014].
- OSUNA ALARCÓN, M. R. & DE LA CRUZ GÓMEZ, E. 2010. Los sistemas de gestión de contenidos en Información y Documentación. *Revista General de Información y Documentación*, 20, 67-100.
- PASTOR SÁNCHEZ, J. A. & MARTÍNEZ MÉNDEZ, F. J. 2009. Aplicación de tesauros, taxonomías y ontologías en los sistemas de gestión de contenidos mediante tecnologías de la Web Semántica. *Ibersid*, 143-153.
- PASTOR SÁNCHEZ, J. A., ORDUÑA MALEA, E. & SAORÍN, T. 2013. Mercado semántico automático en gestores de contenidos: integración y cuantificación. *El profesional de la información*, 22, 381-391.
- PÉREZ MATOS, N. E. & SETIÉN QUESADA, E. 2008. Bibliotecología y Ciencia de la Información: enfoque interdisciplinario. *Acimed*, 18(5), 2-26.
- PÉREZ-MONTORO GUTIÉRREZ, M. 2005. Sistemes de gestió de continguts en la gestió del coneixement. *BiD: textos universitaris de biblioteconomia i documentació*, 14 [Online]. Available: <http://bid.ub.edu/14monto1.htm> [Accessed: 10 de noviembre de 2015].
- PINHEIRO, L. V. R. 2006. Ciência da Informação: desdobramentos disciplinares, interdisciplinaridade e transdisciplinaridade. In: González de Gómez, M.N.; Dill Orico, E. *Políticas de memória e informação*. Natal: EDUFRN, 111-142

- PORTELA, P. 2001. Portales, sindicación, contenidos: nuevas oportunidades para los gestores de información. *El profesional de la información*, 10, 14-16.
- ROBERTSON, J. 2003. *So, what is a CMS?*. [Online]. Available: http://www.steptwo.com.au/papers/kmc_what/ [Accessed 18 de mayo de 2008]
- SÁNCHEZ, J. A. P. 2011. *Tecnologías de la web semántica*, Editorial UOC.
- SANTILLÁN, J. 2007. *Introducción a los Sistemas de Gestión de Contenidos* [Online]. Available: <http://www.slideshare.net/santillan/introduccion-a-los-sistemas-de-gestion-de-contenidos-cms> [Accessed 15 enero 2009].
- SARDUY DOMÍNGUEZ, Y. & URRÁ GONZÁLEZ, P. 2006. Sistemas de gestión de contenidos: en busca de una plataforma ideal. *Acimed*, 14(4) [Online] Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352006000400011&lng=es&tlng=es [Accessed 18 de mayo de 2008].
- STALLMAN, R. 2004. *Software libre para una sociedad libre*, Madrid: Traficantes de Sueños, 2004.
- TAYLOR, R. S. 1966. Professional aspects of information science and technology. *Annual Review of Information Science and Technology*, 1, 15-40.
- TRAMULLAS, J. 2005. Herramientas de software libre para la gestión de contenidos. *Hipertext.net*, 3. [Online] Available: http://www.betsime.disaic.cu/secciones/tec_ja_05.htm [Accessed: 24 de marzo de 2012]
- TRAMULLAS, J. 2009. *Sistemas de Gestión de Contenido* [Online]. Available: <http://www.slideshare.net/tramullas/gestin-de-contenidos> [Accessed 16 mayo 2013].
- WILKOFF, N., WALKER, J., ROOT, N. & DALTON, J. 2001. What's next for content management? *Forrester Research Inc., Cambridge*, 4.

CAPÍTULO II

WEB SEMÁNTICA

CAPÍTULO 2: Web semántica

Se realiza un análisis terminológico, génesis, estructura y aspectos distintivos de la recuperación de la información de la web semántica. Se aborda además su relación con estándares y lenguajes de marcado y ontologías.

2.1. Web semántica: Análisis terminológico

El término Web semántica fue propuesto por primera vez en 1999 por Berners-Lee. Su uso se popularizó a partir de mayo de 2001 con la publicación de uno de los números de Revista American que incluía un artículo relativo a este fenómeno firmado por Berners-Lee, James Hendler y Ora Lassila. Estos autores **definieron a la Web semántica como "una extensión de la Web actual"** exponiendo a discusión abierta la posibilidad de que los ordenadores fuesen capaces de presentar la información contenida en las páginas web y además entender dicha información (Berners-Lee et al., 2001).

De manera general, con la aparición del término se crea un área de estudio prolifera, que propone nuevas técnicas y paradigmas para la representación de la información y el conocimiento; favorecer las posibilidades para compartirla y localizarla, integrarla y recuperar recursos (Castells, 2002b). En los últimos años han comenzado a aparecer aplicaciones basadas en las tecnologías de la Web semántica.

La "Data Web" es el primer paso hacia la completa "Web Semántica", ya que los formatos en que se publica la información en Internet son dispares, como XML, RDF y microformatos. "Data Web" permite un nuevo nivel de integración de datos y aplicación interoperable, haciendo los datos tan accesibles y enlazables como las páginas web (OJEDA, 2014).

Otro ejemplo de ello lo constituye Europeana (ver fig.2.1) cuyo modelo de datos es un nuevo intento de estructurar y representar los datos con los que las diversas instituciones del patrimonio cultural contribuyen a Europeana. El modelo pretende ser más expresivo y flexible. Los principios se basan en buenas prácticas de la Web Semántica y Linked Data, a lo que Europeana quiere contribuir. El modelo en sí se construye sobre estándares establecidos como

RDF(s), OAI-ORE, SKOS y Dublin Core. Actúa como una ontología común de alto nivel que permite la interoperabilidad al mismo tiempo que mantiene los modelos de datos originales y sus perspectivas de información (Doerr et al., 2010).

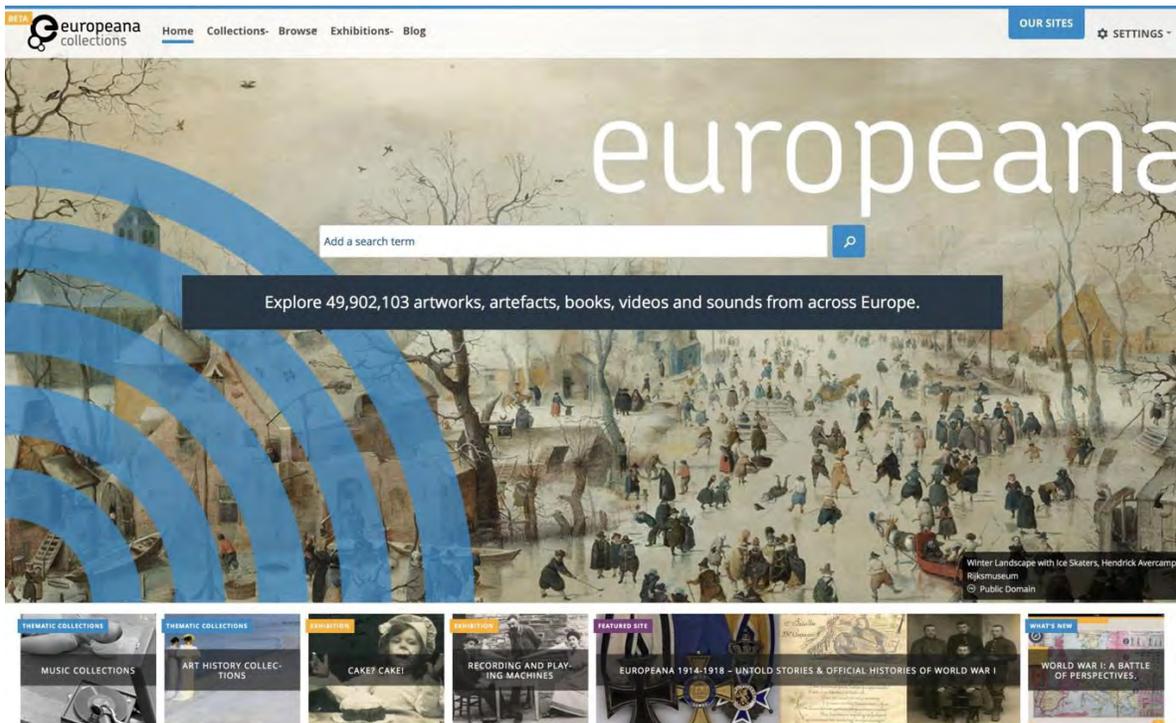


Fig. 2.1. Ejemplo de implementación de Web Semántica: Visualización de la Página principal de Europeana (Europeana, 2016)

A partir de experiencias acumuladas, tanto teóricas como prácticas, varios autores han definido a la Web semántica desde diversas perspectivas, que van desde su contextualización en la Inteligencia Artificial hasta áreas menos tecnológicamente equipadas y más enfocadas al desarrollo social.

Desde la visión de las Ciencias de la Información, Rodríguez Perojo y Ronda León, cuando definen la Web semántica asumen, al igual que sus creadores, **que es "una extensión del web actual"** (Berners-Lee et al., 2001) pero analizada como herramienta para el intercambio y consulta de información su "idea básica es tener los datos definidos y relacionados para que su uso sea más efectivo y sea posible su automatización, integración y re-utilización por medio de diferentes aplicaciones, es decir, pretende proporcionar una

infraestructura que permita que las páginas web, las bases de datos, los programas y aplicaciones, los dispositivos, tanto personales como los empleados en el hogar, puedan consumir y producir datos, sin los problemas causados por los diferentes protocolos de acceso a la información que hacen de la transferencia de contenidos una tarea **ardua y difícil**” (Rodríguez Perojo and Ronda León, 2005). Es un concepto que particulariza las acciones metainformativas de usuarios y/o productores de la información. Consecuentemente, proponen un mapa conceptual para desglosar sus principales características y apoyar la conceptualización del término (Ver fig. 2.2)

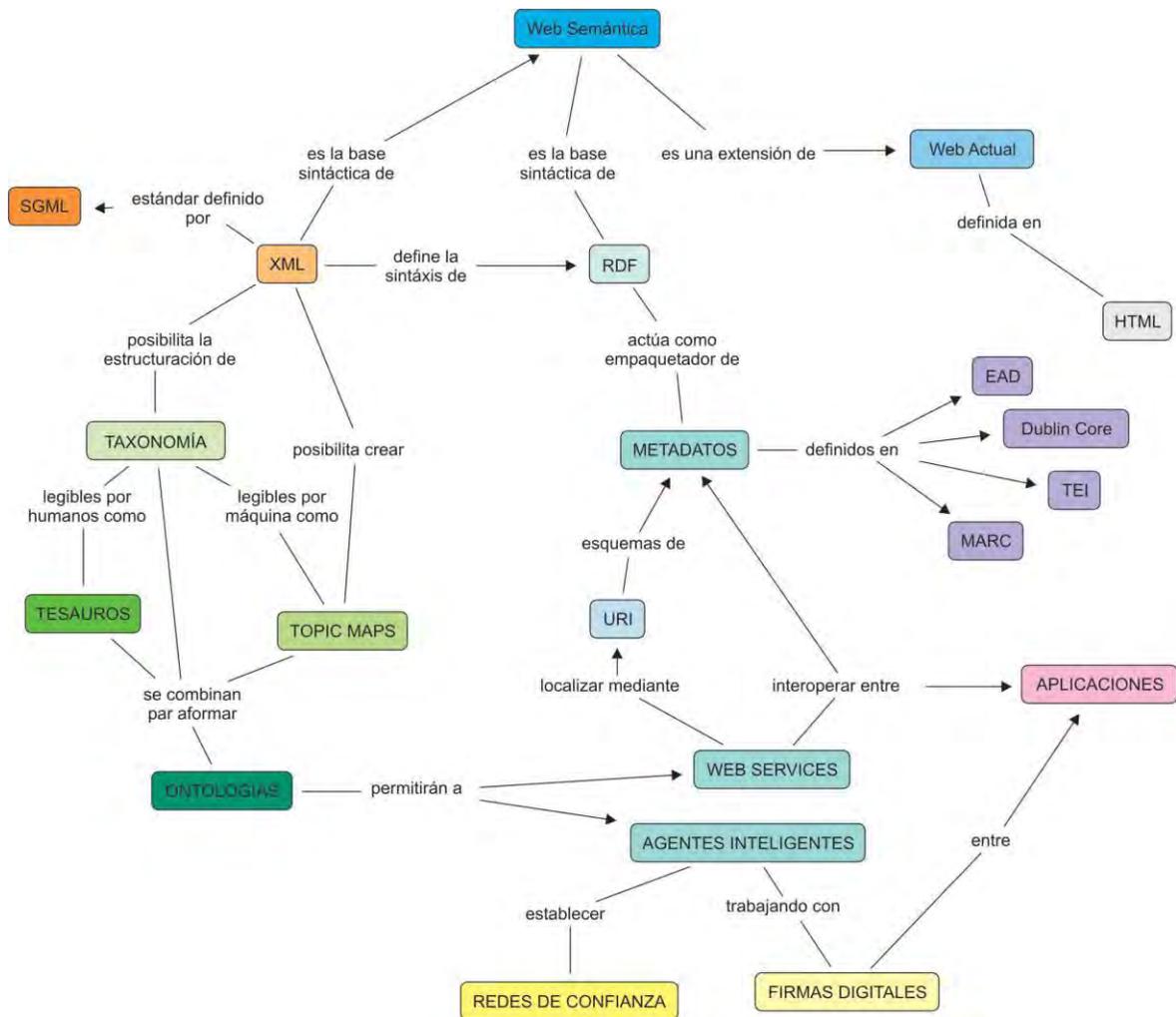


Fig. 2.2. Mapa conceptual del término Web Semántica (Rodríguez Perojo and Ronda León, 2005)

Para Codina y Rovira, la Web semántica es el proyecto del W3C para transformar la Web en la Web de las próximas décadas. En 2006 apuntaban sobre dos aspectos que evidenciaban el paso de su desarrollo al distinguirla a manera de una auténtica idea-fuerza, capaz de movilizar energías en función de su avance. En segundo lugar, aportando nuevos estándares que ya son de uso habitual (como el lenguaje XML) e influenciando en el desarrollo de la nueva generación de navegadores y editores de páginas web (Codina Bonilla and Rovira, 2006).

Estos autores reconocen dos vertientes fundamentales que pautan a su entender el concepto de Web semántica: desde la visión de la inteligencia artificial (basada en la posibilidad y capacidad de un (o varios) software(s) de simular el razonamiento humano) y desde la visión del procesamiento robusto de datos (basado en el tratamiento sistémico de los documentos y la asignación de metadatos) (Codina Bonilla and Rovira, 2006). Así proponen dos definiciones:

Definición 1. La visión de la Inteligencia Artificial: La Web semántica es un conjunto de iniciativas destinadas a promover una futura Web cuyas páginas estén organizadas, estructuradas y codificadas de tal manera que los ordenadores sean capaces de efectuar inferencias y razonar a partir de sus contenidos (Codina Bonilla and Rovira, 2006).

Definición 2. La visión del procesamiento robusto: La Web semántica es un conjunto de iniciativas destinadas a convertir la World Wide Web en una gran base de datos capaz de soportar un procesamiento sistemático y consistente de la información (Codina Bonilla and Rovira, 2006).

Para García la Web semántica es básicamente “un área pujante en la confluencia de la Inteligencia artificial y las tecnologías Web que propone introducir descripciones explícitas sobre el significado de los recursos, para permitir que las propias máquinas tengan un nivel de comprensión de la Web suficiente como para hacerse cargo de una parte, la más costosa, rutinaria, o físicamente inabarcable, del trabajo que actualmente realizan manualmente los usuarios que interactúan en la Web” (García Ruiz, 2008).

El proyecto W3C, reconocido como el organismo que regula aspectos esenciales de la Web, expone que la Web semántica proporciona un marco común que permite que los datos sean compartidos y reutilizados a través de aplicaciones, empresas y fronteras comunitarias. Es un esfuerzo colaborativo liderado por el W3C con la participación de un gran número de investigadores y socios industriales. Está basado en Resource Description Framework (RDF) e integra una variedad de aplicaciones utilizando XML para la sintaxis y URI para las denominaciones (W3C, 2013).

Por otra parte, lingüísticamente hablando, el término semántica implica, a decir de Becerra Ocampo (2013) el reconocimiento de semejanza de significantes, polisemia, sinonimia, concepciones de la realidad y sus diferentes expresiones lingüísticas, etc. (ver fig. 2.3)

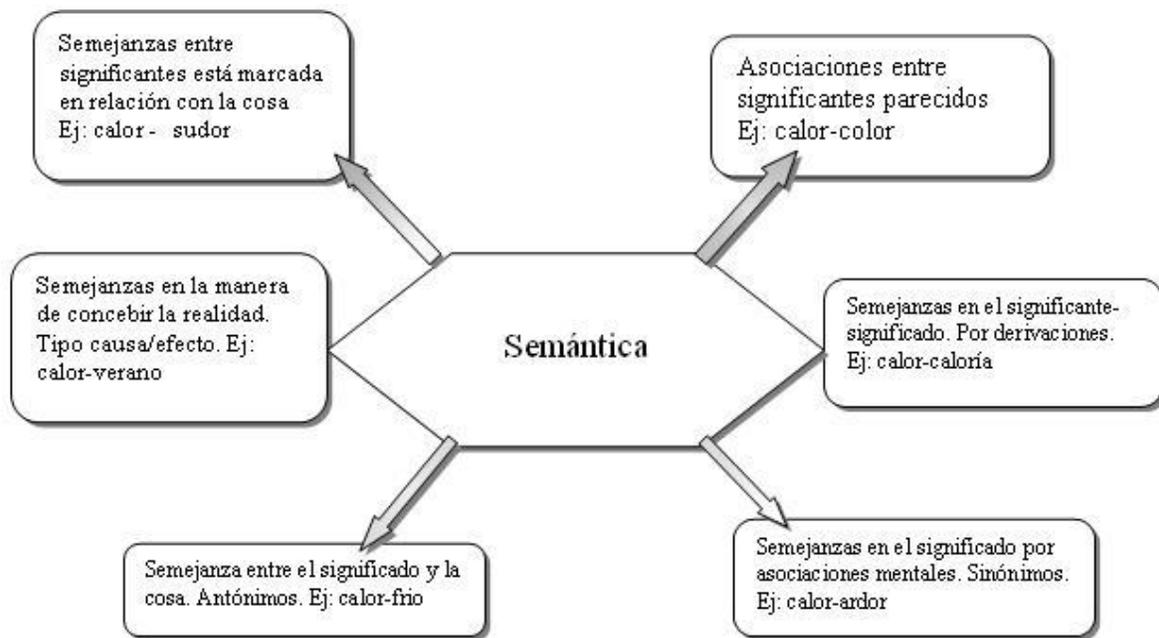


Fig. 2.3. Implicaciones del término semántica (Becerra Ocampo, 2013)

Por tanto, dotar a la web de semántica, es crear mecanismos y herramientas que reconozcan en la información digital existente en la web, registrada en diferentes formatos, las diferentes variantes y la riqueza del lenguaje natural, al permitir organizar y recuperar dicha información tomando en consideración

su significado comunicacional y no solo la estructura sintáctica de los términos que la identifican.

2.2. Origen e historia de la Web semántica

Internet ha revolucionado en muy poco tiempo el modo de comunicarnos y de acceder a la información. En 1989 Tim Berners-Lee, informático de la Organización Europea para la Investigación Nuclear (CERN, por sus siglas en inglés), en Ginebra (Suiza), crea la World Wide Web (WWW, por sus siglas en inglés), llegando a estandarizar los protocolos de conexión TCP/IP. Interesado en mejorar la gestión de la información acerca de los aceleradores y experimentos que se hacían en el CERN, propuso una solución basada en un sistema distribuido de hipertextos que pretendía resolver el problema de la pérdida de la información y permitir el intercambio de información entre los científicos de dicho centro de investigación (Hidalgo Delgado and Rodríguez Puente, 2013).

A medida que ha ido avanzando la tecnología a través de los años, la web desde su invención también ha tenido cambios considerables (Ver Fig. 2.3). Castells , Perojo y Ronda, coinciden en afirmar que, a partir del surgimiento de Internet, sucedieron un conjunto de acontecimientos que tributaron a la evolución de la Web (Castells, 2002a, Rodríguez Perojo and Ronda León, 2005):

- 1989: Tim Berners Lee presenta su proyecto WWW en el CERN (Conseil Européen pour la Recherche Nucléaire).
- 1993: Creación de los primeros servidores Web y el navegador Mosaic¹⁸.
- 1994: Creación del Consorcio Web (World Wide Web Consortium o W3C).
- 1997: Creación de SHOE (Simple HTML Ontology Extensions), primer antecedente de la web semántica, basado en HTML¹⁹.

¹⁸ En enero de 1993 eran conocidos 50 servidores HTTP en todo el mundo, cifra que se elevó a 200 en tan solo diez meses. Este fue el inicio de la explosión de la web, al punto de que a finales de 1994 existían más de 10.000 servidores y 10 millones de usuarios en el mundo

Antes de 1990, Internet era un archipiélago de ordenadores inconexos. Tampoco existían los buscadores, ni se podían integrar imágenes en la pantalla puesto que no habían nacido las interfaces gráficas. Como se presentó en el capítulo anterior, las primeras páginas Web que se publicaron en forma de páginas HTML editadas manualmente. Luego, enlazaban unas páginas con otras. Estas primeras versiones de la web se conocen en la literatura como web 1.0 y se extendió alrededor de los años 1990 al 2000 (Ver Fig. 2.4).

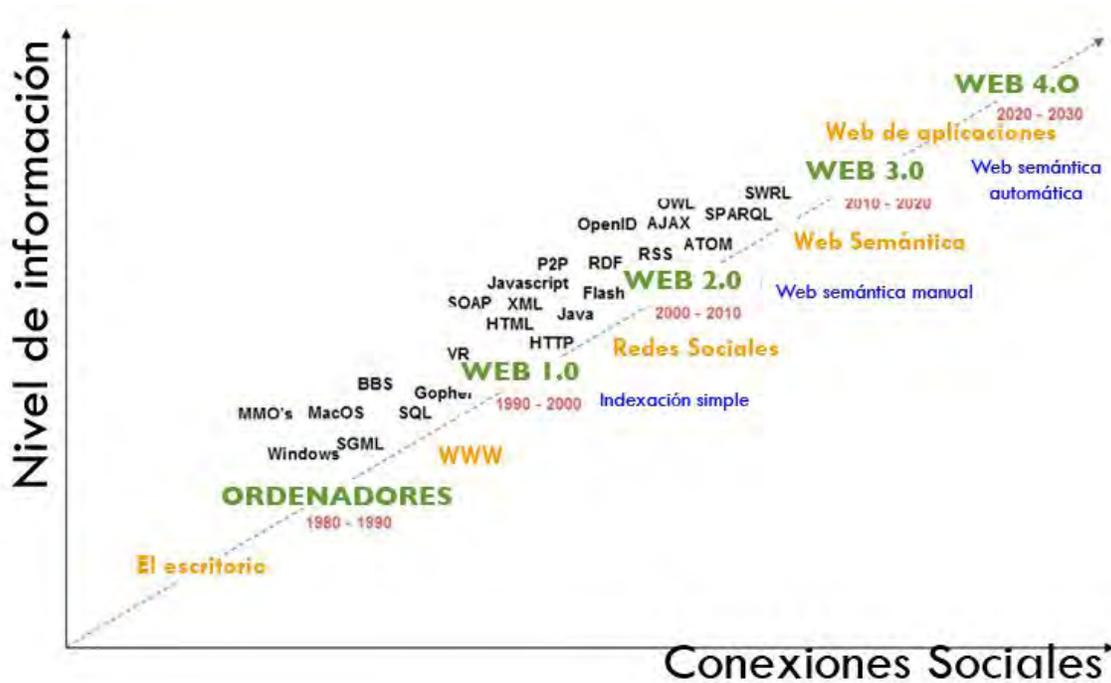


Fig. 2.4. Evolución de la Web (Garzón, 2015)

La Web 1.0, aunque surge como idea en los años 60, marca su inicio a finales de los 80' y principio de los 90' por su creador Tim Berners-Lee. Constituye el punto de apertura en el desarrollo de las telecomunicaciones (Ver fig. 2.4 y 2.6). Fue un hecho importante dadas las posibilidades de conectividad entre las máquinas y el crecimiento exponencial de la cantidad de información que

¹⁹ En 1997, la cifra se elevó a más de 650.000 servidores web y en octubre del 2011, según las estadísticas publicadas por la Compañía de Servicios de Internet Netcraft, existían en el mundo 448.164.344 de estos servidores con aproximadamente 2.095.006.005 usuarios de Internet

ofrecía. Las páginas eran estáticas, es decir, destinadas a la lectura, meramente informativas y trabajadas en HTML.

Se distinguió por el uso de navegadores de solo texto como ELISA, bastante rápidos, después surgió el HTML que hizo las páginas web más agradables a la vista y los primeros navegadores visuales como IE, Netscape, etc. (Muñoz, 2015), la explotación de bases de datos y de lenguajes como SQL para su **construcción, el surgimiento de los "groupware"** (Spivack, 2007), pero lo más relevantes de la etapa fue la aparición del World Wide Web (conocida como la **WWW**). Sin embargo, aunque no existían posibilidades de retroalimentación o desarrollo de servicios en línea que dependiesen de la interacción con el usuario, con ella, las grandes empresas inauguraron su estrategia online. Fue usada fundamentalmente para publicar documentos y realizar transacciones (OJEDA, 2014). Para esta etapa la búsqueda y recuperación de la información era precaria, basada en una indexación simple.

En general, el término que ha sido creado para describir la Web antes del impacto de la fiebre "punto com" en el 2001, que es visto por muchos como el momento en que el internet dio un giro (OJEDA, 2014).

La Web 2.0 (Ver fig. 2.4), **término creado por O'Reilly en 2004, aportó** significativos avances en la tecnología Web. Con respecto a la 1.0, constituyó un cambio radical en los procesos de producción e intercambio de información y colaboración entre usuarios.

Hasta ese momento, solo unos pocos usuarios tenían los conocimientos necesarios para crear contenido que se mostrara a través de la Red de redes. Con la Web 2.0, el contenido estático dejó paso a una web dinámica en la que los usuarios, que habían sido limitados hasta el momento a la observación pasiva de la información que se generaba para ellos, conseguían aportar sus propios contenidos de manera sencilla (Evans, 2011).

Surgen redes sociales como Facebook y twitter, los servicios web, servicios de red social, las wikis, blogs, folksonomias, entre otros. Germinan comunidades virtuales con usuarios como generadores del contenido publicado. Todo ello es consecuencia del principio de colaboración sobre el que se construye y

desarrolla la Web 2.0, donde los usuarios se convierten en contribuidores que publican informaciones y realizan cambios en los datos. Como resultado, además de las redes sociales, aparecen ecosistemas informativos como Google²⁰, Wikipedia²¹, Ebay²², YouTube²³, Skype²⁴, Writely²⁵, Blogger²⁶, entre otras, donde existe una proporcionalidad directa entre el número de usuarios y el valor del sitio y sus contenidos. A la vez, asoman importantes cambios en los códigos que van a permitir la introducción de palabras claves para la recuperación de información en los motores de búsqueda utilizando algoritmos complejos (Keyword search). En esta etapa continúa desarrollándose los códigos HTML y aparecen otros como XML, SOAP y Javascript (Garzón, 2015, Spivack, 2007).

Además emergen los CMS como facilitadores en la construcción de páginas y sitios web dinámicos.

Con la transición hacia la Web 3.0 comienza a trabajarse en la semantización de la Web, bajo el concepto de la Web semántica entendida y asumida como reto. Se recopilan mecanismos de las Web anteriores y se optimiza a través de un rápido y fácil acceso al intercambio de información y la participación en redes sociales, facilitando las actividades del usuario. Para ello, y como

²⁰ Es más que un buscador, es una marca que abarca desde un sistema de correo electrónico (Gmail) hasta un mapa mundial (Google maps). OJEDA, H. 2014. Mapa Conceptual Evolución de la Web. *Diplomado Componente Docente en Educación Interactiva a Distancia/ Módulo III: Estrategias, Herramientas y Recursos Educativos*. Barquisimeto: Universidad Fermín Toro.

²¹ Es un proyecto de enciclopedia libre donde son los usuarios los que escriben los artículos. Ibid.

²² Es un sitio que permite a millones de usuarios en todo el mundo comprar y vender cualquier cosa en cualquier sitio. Ibid.

²³ Es el sitio favorito de miles de usuarios para ver vídeos que los mismos usuarios aportan y comparten. Ibid.

²⁴ Es una aplicación que permite hablar gratuitamente con otros usuarios que se encuentran en cualquier parte del mundo, usando tecnología de voz sobre IP. Ibid.

²⁵ Ofrece un procesador de texto en línea que permite compartir los documento con otros usuarios e incluso publicar automáticamente en Blogger. Ibid.

²⁶ Es la aplicación más popular para crear un weblog de forma fácil. Ibid.

característica distintiva, se trabajan proyectos de asignación de metadatos semánticos y ontológicos a la WWW, como informaciones adicionales al texto primario, para describir el contenido, significado y establecer relación entre los datos, para que sea posible su entendimiento y evaluación automática por máquinas de procesamiento (inteligencia artificial).

Tanto la Web 1.0 como la 2.0 forman parte de la llamada Web sintáctica, donde puede hablarse de una colección de datos desestructurados, esencialmente en la Web 2.0 se enlazan los documentos (ver Fig.2.5) pero, el sistema de codificación de las páginas carece de interpretación semántica certera y aún más por parte de analizadores automáticos. Por tanto, las páginas web tienen semántica únicamente para los humanos.

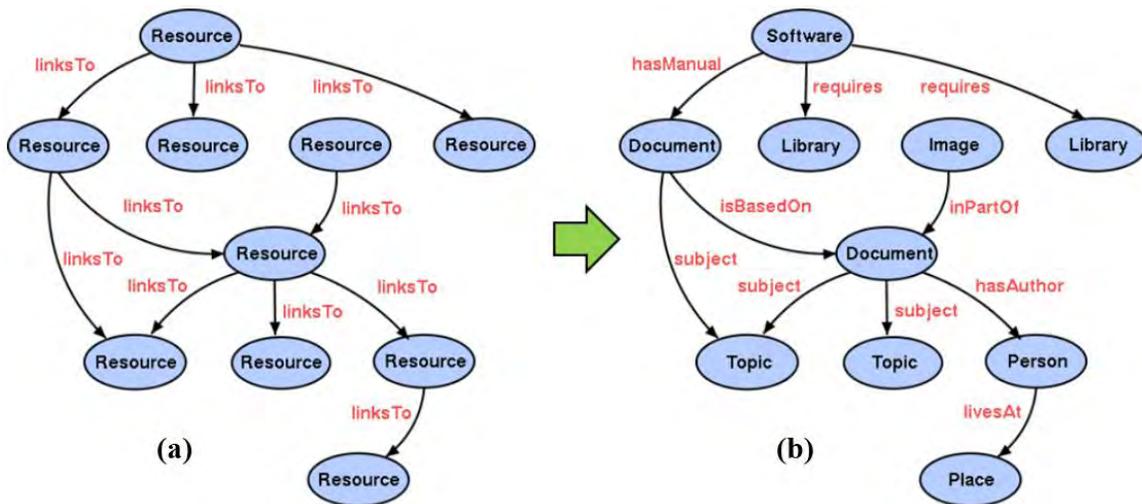


Fig. 2.5. Diferencia gráfica entre la Web sintáctica (a) y la Web semántica (b) (fuente: <https://www.analistaseo.es/wp-content/uploads/2014/09/web-actual-web-semantica.png>)

La idea es integrar en la web actual, datos relacionados entre ellos, bien semánticamente o por atributos que los determinan (ver fig.2.5). Para lograr una web semántica los datos de las páginas se deberán estructurar y escribir con un único formato que permita a las máquinas seleccionar mejor la información que exista en Internet. La web semántica viene a ser como una filosofía creada para el intercambio de datos cuya información es leída e interpretada entre máquinas, facilitando la comunicación entre el ser humano y un sistema informático (Pastor Sánchez, 2011b). En esta nueva web, un objeto

será descrito de la misma forma, por tanto, los buscadores podrán identificarlo y ligarlo con posibles búsquedas relacionadas.

La Web 3.0 no constituye un reemplazo de la web actual, sino que debe ser vista como una adición de contenido a la web. La idea es gestionar un conjunto de datos y metadatos, el todo y sus partes de manera holística y semántica, que permita el funcionamiento sobre la web de manera automática. Con ello se podrá describir sintáctica y semántica de los documentos.

En esta etapa se utilizan lenguajes como RDF, OWL, SPARQL, RSS; sistemas Open ID, bases de datos y búsquedas semánticas, entre otras. En la actualidad, está en pleno desarrollo y, desde un punto de vista técnico, los pilares fundamentales en que se basa son las ontologías y las etiquetas.

Con la web 3.0 mejora Internet desde la interoperabilidad entre los sistemas informáticos, con agentes inteligentes que utilizan datos semánticos y descripciones lógicas, utilizando reglas que expresan relaciones entre conceptos y datos en la red, así como aumento de la integración de fuentes de datos heterogéneas.

Algunos autores (Yu, 2011, Kambil, 2008, Benito-Osorio et al., 2013, OJEDA, 2014), desde una visión previsor de lo que podrá ser la evolución de la web en los próximos años, han presentado y definido desde la Web 4.0 o Meta Web basada en conexiones inteligentes (ver fig. 2.4 y 2.6), hasta la Web 7.0. (Ver fig. 2.6)

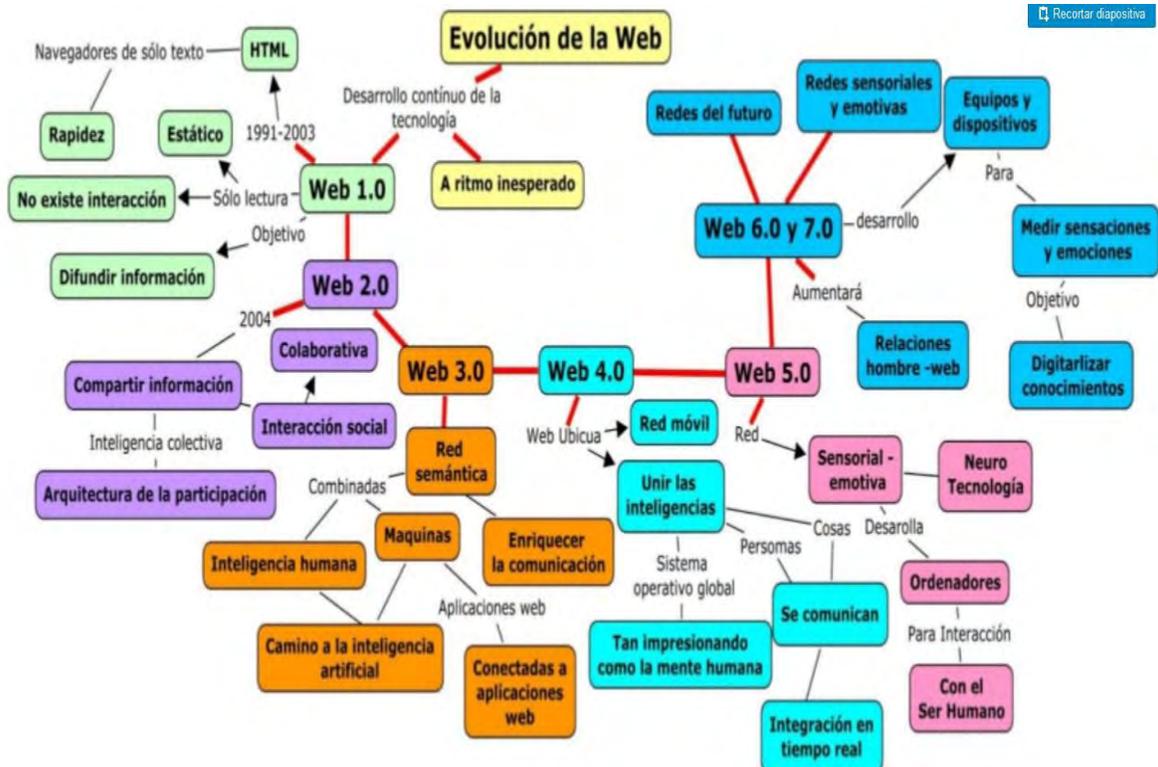


Fig. 2.6. Evolución de la Web: DESDE LA Web 1.0 hasta la Web 7.0 (Fuente: Olivera, J. (2013) citado por (OJEDA, 2014))

2.3. Estructura de la Web semántica

La web actual cuenta con un conjunto de datos, que corresponden a lo primero que se perseguía: la visualización del documento (lo que se consiguió con HTML) y no a la identificación del contenido del mismo. Por tanto, los buscadores (browser) se concentraron en sus inicios en localizar documentos. A esos datos, con la filosofía de la web semántica se le fueron vinculando otros con nuevos formatos de representación semántica básica como los RDF, que actualmente constituyen la base de la web semántica. (Ver fig.2.7 y 2.8)

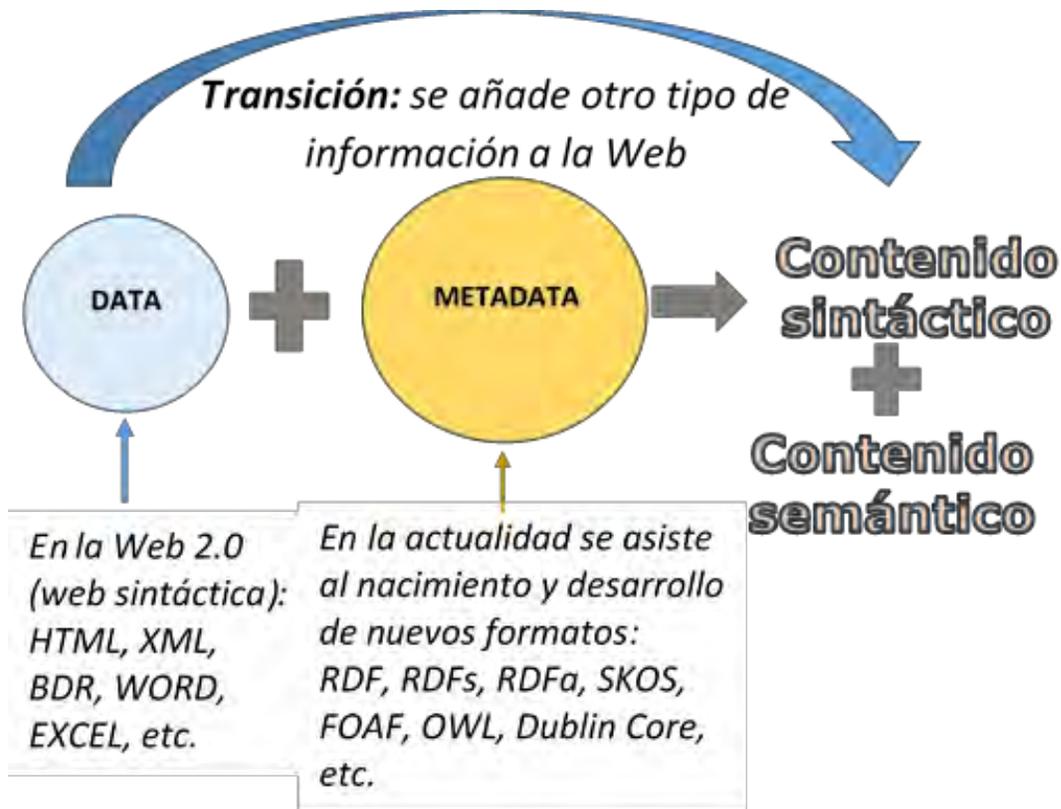


Fig. 2.7. Modelización de la transición de la web actual a la web semántica (elaboración propia)

Los principales componentes de la Web Semántica son los metalenguajes y los estándares de representación XML, XML Schema, RDF, RDF Schema, OWL, así como el lenguaje SPARQL para la consulta de datos RDF (Ver fig. 2.8):

- XML aporta la sintaxis superficial para los documentos estructurados, pero sin dotarles de ninguna restricción sobre el significado.
- XML Schema es un lenguaje para definir la estructura de los documentos XML.
- RDF es un modelo de datos para los recursos y las relaciones que se puedan establecer entre ellos. Aporta una semántica básica para este modelo de datos que puede representarse mediante XML.
- RDF Schema es un vocabulario para describir las propiedades y las clases de los recursos RDF, con una semántica para establecer jerarquías de generalización entre dichas propiedades y clases.

- OWL es un lenguaje para definir ontologías mediante la descripción detallada de propiedades y clases: tales como relaciones entre clases, cardinalidad, igualdad, tipologías de propiedades más complejas, caracterización de propiedades o clases enumeradas.
- SPARQL es un lenguaje de consulta de conjuntos de datos RDF. Además en dicha especificación también se incluye un formato XML que detalla el modo en el que se estructuran los resultados obtenidos.

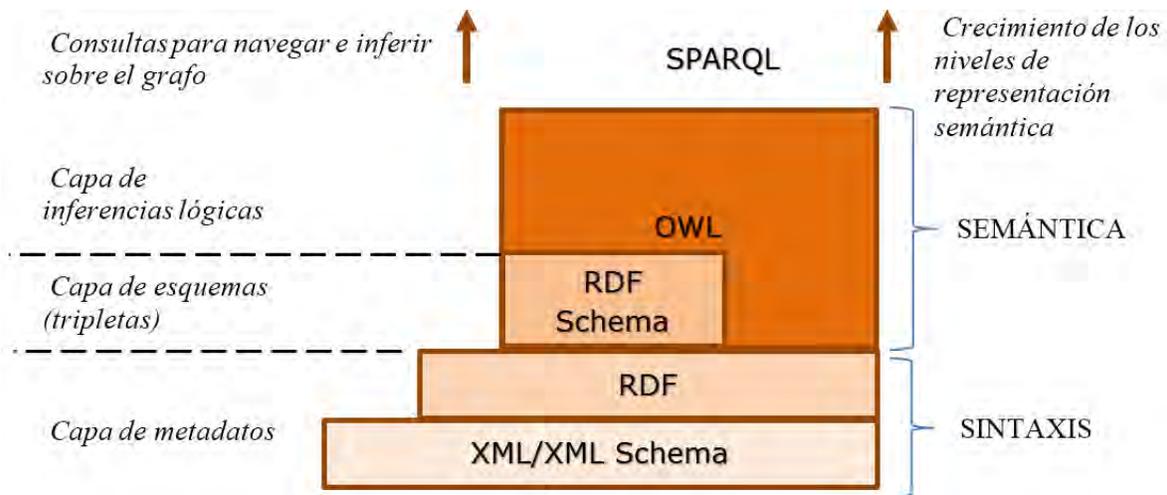


Fig. 2.8. Base semántica de la Web semántica (elaboración propia)

Sobre esta premisa se han sentado las bases de los diferentes modelos de representación de la estructura de la web semántica defendido por diversos autores (Berners-Lee, 2000, Codina Bonilla and Rovira, 2006, Sánchez, 2011, Rodríguez Perojo and Ronda León, 2005). Hasta el momento, tras conclusión empírica de la autora después de haber consultado varias fuentes, uno de los modelos más citados es el de Tim Berners Lee, James Hendler y Ora Lassila, dado a la luz en mayo de 2001, donde explican de forma sencilla su idea de la Web Semántica y los primeros pasos que hay que dar para llevarla a cabo. Según dichos autores, la arquitectura de la Web Semántica se podría representar de la siguiente forma: (ver fig.2.9)

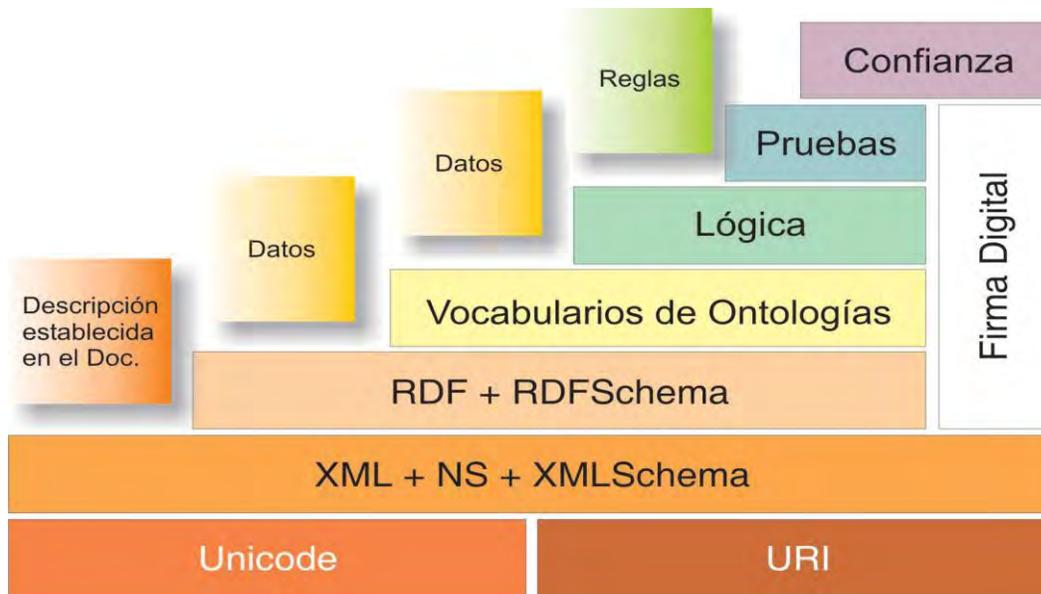


Fig. 2.9. Modelo de estructura en capas de la Web semántica (Berners-Lee, 2000)

Este diagrama, presentado por Berners-Lee en la XML Conferencia del año 2000, ha servido como aproximación visual al conjunto de tecnologías que forman el esquema de capas mencionado, cuyos componentes son:

- En la base quedan representados los estándares para la localización de recursos de información en el web de forma inequívoca y única como son los URIs y la norma internacional Unicode para la codificación de caracteres a nivel internacional.
- XML como base sintáctica para la estructuración del contenido en el web, así como el empleo de espacios de nombres (Namespaces) para asociar con precisión cada propiedad con el esquema que define dicha propiedad y esquemas (XML Schema) para definir qué elementos debe contener un documento XML, cómo están organizados, qué atributos y de qué tipo pueden tener sus elementos.
- Un modelo básico para establecer propiedades sobre los recursos, para el que se empleará RDF, así como un modelo para definir relaciones entre los recursos por medio de clases y objetos, que se expresan mediante esquemas en RDF (RDF Schema).

- Lenguajes para la representación de ontologías que permitan la interoperabilidad y reutilización entre ontologías de diversos dominios del conocimiento en el web, cuya base se encuentra en RDF Schema.
- Una capa lógica que permita realizar consultas e inferir conocimiento, donde estarían las ontologías, agentes software y web services como estructuras para lograr interoperabilidad entre aplicaciones y sistemas de información heterogéneos.
- Una capa de seguridad que permita asignar niveles de fiabilidad a determinados recursos, de forma comprobable posteriormente por los agentes, para lo que se utilizarán firmas digitales y redes de confianza.
- Dicho modelo, es la propuesta de desarrollo de la Web semántica ofrecida por el W3C

Con el transcurso del tiempo han aparecido representaciones más complejas, como la que exhibe la Wikipedia (ver fig.2.10), pero, generalmente, todas siguen los mismos componentes básicos: XML para la sintaxis, XML Schema para soportar la estructura de los documentos, RDF para modelar los datos y sus relaciones aportando una semántica básica, RDF Schema para la descripción de propiedades y clases de recursos y establecer jerarquías de generalización, OWL como lenguaje para definir ontologías y SPARQL para las consultas. Pues estos son los lenguajes básicos que sustentan la estructura de la Web semántica.

información de su interés” (García-García, 2014). Autores como Mai (2004) y García-García (2014) han reconocido que la clasificación bibliográfica puede aportar a la organización de los contenidos de la Web, si se ajusta a los requerimientos del medio digital. Sin embargo, sólo en entornos controlados se ha procedido a sistematizar el uso de esquemas de clasificación, como en bibliotecas digitales, repositorios, o entornos corporativos (García-García, 2014).

La anotación semántica de los documentos es el primer paso para permitir el procesamiento automático de la información de la Web, y por tanto para la creación de la Web Semántica. Las anotaciones semánticas llevan implícito un proceso de análisis, extracción y marcado de la información para enriquecerla semánticamente (Navarro Galindo, 2012).

Pueden identificarse diferentes tipos de anotación. Entre las más populares se encuentra la tipología establecida a partir del agente que realiza la anotación: automática o asistida (también conocida como anotación manual o directa por autores como Navarro Galindo) seguida por autores como Pastor Sánchez y Navarro Galindo (Pastor Sánchez et al., 2013, Navarro Galindo, 2012).

En la anotación automática es realizada por el software. Parte de un proceso automático o semi-automático. El objetivo final es asignar a cada elemento del texto una etiqueta correcta y única, eliminando, en primer lugar cualquier rastro de ambigüedad y con ello la necesidad de la participación humana en el proceso. Se identifican las entidades semánticas y sus relaciones en el contenido fuente y se realizan correspondencias o mapeos con las entidades y relaciones de ontologías (Pastor Sánchez et al., 2013, Navarro Galindo, 2012).

En la anotación asistida, un anotador o revisor humano es el encargado de determinar la etiqueta final de aquellos elementos que el sistema informático no haya sido capaz de anotar bien o unívocamente, aún después de ejecutar las rutinas de desambiguación adecuadas. Es decir, el usuario es el responsable de analizar los documentos, identificar las entidades semánticas y las relaciones entre ellas (Pastor Sánchez et al., 2013, Navarro Galindo, 2012).

Otra manera de clasificarla es de acuerdo al lugar donde se almacenan las anotaciones. Pueden ser internas (o embebidas) o anotaciones externas. Esta clasificación es seguida por autores como Popov y Navarro Galindo . No son excluyentes dado que existe la posibilidad de usar un almacenamiento dual de anotaciones, dependiendo del sistema que se pretenda desarrollar (Popov et al., 2004, Navarro Galindo, 2012).

Las anotaciones internas o embebidas son creadas con lenguajes de marcado, por ejemplo RDFa, y almacenadas dentro del mismo documento dónde se realiza la anotación. Las anotaciones externas, como el nombre sugiere, son almacenadas en ficheros o en servicios distintos del documento que se anota y se accede a ellas mediante lenguajes como XML, OWL o RDF (Navarro Galindo, 2012).

En el contexto de los S-CMS la posibilidad de realización de anotación semántica de los contenidos viene dada por el uso de lenguajes de marcado y/u ontológicos: XML, RDF/ RDFa, OWL. Con ellos se establecen las marcas semánticas y las reglas sintácticas que se utilizan para describir, desde el software, la estructura del documento digital y representar su significado. (Ver capítulo 3 sobre CMS semánticos).

2.5. Estándares y lenguajes de marcado

La mayoría de la información que se encuentra actualmente en Internet es textual y sin ninguna estructura formal que permita a las máquinas entender y aprovechar dicha información de manera automática y con ella trabaja el CMS tradicional. Sin embargo, en una red social todos los contenidos que están compartiendo son valiosos, por lo tanto es necesario definir mecanismos para dar estructura y enriquecer semánticamente la información existente, para que esta sea aprovechada por otras aplicaciones software (Caldón et al., 2010). Tanto para la realización de técnicas de procesamiento de lenguaje natural, de anotaciones semánticas en sus distintas variantes, o de cualquier tecnología semántica como los S-CMS, requieren de estándares, lenguajes de marcado especializados para lograr representar, identificar las estructuras en los documentos, agrupar y relacionar los conceptos de un dominio de

conocimiento, de la manera menos ambigua posible, facilitando su recuperación.

El primer exponente a partir del cual se traza el surgimiento de los lenguajes de marcado, viene dado con la aparición de SGML (Standard Generalized Markup Language) que es una norma de estandarización derivada de GML (Generalized Markup Language), la cual define que el etiquetado debe describir la estructura del documento electrónico y otros atributos de forma que puedan ser interpretados y procesados por una máquina (García Ruiz, 2008). Así mediante marcas se delimitaban y describían los componentes del documento electrónicos

Ante el reto de la organización y recuperación de información electrónica, los lenguajes de marcado se han convertido en un nuevo alfabeto semántico para los productores y para los gestores de información (Méndez Rodríguez, 1999).

En los últimos años se han creado numerosas aplicaciones en pos de mejorar las condiciones representacionales y de recuperación de información en la Web. Los principales medios con los cuales se persiguen los objetivos de la web semántica son: mediante una codificación de páginas en la cual las etiquetas transporten carga semántica (XML) y aportando descripciones (metadatos) de las páginas y sitios web con un formato que sea compatible con la estructura general de la WWW y con diversas categorías de páginas, e interoperable entre distintos sistemas informáticos (RDF). A todo ello se integra un sistema de ontologías que permita especificar conceptos de diversos dominios mediante el uso de un lenguaje basado en lógica simbólica y susceptible a ser interpretado por un ordenador (OWL) (Pedraza Jiménez et al., 2007).

En los sub-epígrafes siguientes se abordarán dichos estándares para el entorno de la Web Semántica bajo una visión general de las características que aporta cada uno, XML, RDF y OWL, tomando en consideración que sobre los mismos

se enriquecen las estrategias de trabajo en la construcción de la Web semántica.

2.5.1. XML: Extensible Markup Language

XML son las siglas utilizadas para el término anglosajón *Extensible Markup Language* que en español se traduce como Lenguaje de Marcado Extensible. Es una especificación que define una forma estándar de añadir información adicional (marcas) a los documentos y como lenguaje de marcado permite identificar las estructuras en un documento (Walsh, 1998).

XML es un formato sencillo, basado en texto para representar información estructurada: documentos, datos, configuración, transacciones, facturas, entre otras (W3C, 2015). Fue derivado de un formato estándar antiguo llamado SGML²⁸ (ISO 8879). Surge tras el interés por la evolución del web hacia sistemas flexibles y mejor estructurados según su contenido, pues las alternativas viables (HTML y SGML) no eran prácticas para este propósito (Walsh, 1998). Indujo la necesidad de una restricción estructural para proporcionar métodos inequívocos de expresión semántica (Méndez Rodríguez, 1999).

XML es muy parecido a HTML, excepto que en XML se pueden definir etiquetas propias. Esta capacidad de producir documentos personalizados es muy útil cuando se necesita para realizar un seguimiento de algunos tipos de datos, sobre todo en el mundo de la web (específicamente en la edición y el comercio electrónico), lo que le permite hacer una "mini-base de datos" de cada

²⁸SGML es el Lenguaje de marcado generalizado estándar definido por la norma ISO 8879. SGML ha sido el estándar de manera independiente del proveedor para mantener repositorios de documentación estructurada desde hace más de una década, pero no es muy adecuado para servir a los documentos a través de Internet (por un número razones técnicas más allá del alcance de este artículo). Definición de XML como un perfil de aplicación de SGML significa que cualquier sistema de SGML totalmente conforme será capaz de leer documentos XML. Sin embargo, usar y entender documentos XML no requiere un sistema que es capaz de entender la generalidad completa de SGML. XML es, en términos generales, una forma restringida de SGML (WALSH, N. 1998. What is XML. *XML. commune.*).

documento sin la sobrecarga y los gastos que muchas bases de datos cargan a un proyecto Web (Myer, 2003).

Las reglas de sintaxis de XML son estrictas (sus componentes se muestran en la Fig.2.11), pues no procesan archivos que contengan errores, sin embargo advierte con mensajes de error para solucionarlos. Por tanto, casi todos los documentos XML pueden ser procesados de forma fiable por el software del ordenador. En el sitio W3C, XML se reconoce como uno de los formatos que actualmente más se utiliza para el intercambio de información estructurada: entre programas, equipos y personas, tanto a nivel local como a través de redes (W3C, 2015).

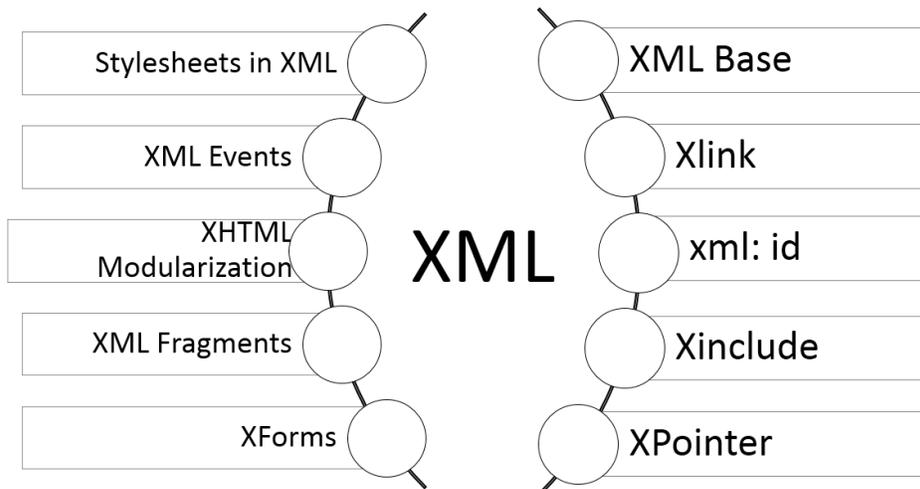


Fig. 2.11. Componentes de XML según W3C (elaboración propia)

Dada sus características y potencialidades, XML es la capa sintáctica base de la web semántica. Todas las otras tecnologías que proporcionan características de la web semántica se construirán sobre la base de XML (Daconta et al., 2003). En un CMS contribuye esencialmente como lenguaje de descripción y además es muy útil para la migración de datos (Liendl et al., 2001). Son varios los CMS que se implementan con este lenguaje. Entre los más conocidos puede citarse a WordPress, Joomla, Drupal.

A decir de algunos autores (Pedraza-Jimenez et al., 2007), XML es el elemento de la web semántica que con mayor fuerza a irrumpido en el contexto profesional de la bibliotecología y la documentación.

Sobre la base de XML, se definen distintos lenguajes de marca para los diferentes tipos de documentos. Según Márquez Solís (2007), entre los más conocidos se encuentran (Márquez Solís, 2007):

- SHOE: lenguaje de representación del conocimiento basado en HTML (extensión de las etiquetas XML a las que se añade significado)
- OWL (Online Writing Lab): habilita el procesado semántico de la información mediante máquinas.
- OIL: lenguaje de intercambio de ontologías. La unión de DAML+OIL puede derivar en la vía más práctica para que los agentes inteligentes comprendan el contenido de las páginas y, por tanto, desarrollen la Web semántica.
- XOL: es un lenguaje para el intercambio de ontologías. Se usa como lenguaje intermedio; maneja datos, definiciones de clases y de objetos, bases de hechos, etc. XOL tiene una sintaxis basada en XML y una semántica que emula el modelo de conocimiento.
- OKBC (Open Knowledge base Connectivity): Es expresivo, fácilmente legible, comprensible y procesable por agentes humanos y no humanos.
- IFF: es el lenguaje de marcas asociado a la Information Flow Framework. Extiende XML incorporando ideas de RDF, OIL, OML.
- RDF (Resource Description Framework): es una recomendación del W3C, basado en XML, que proporciona la tecnología para describir metadatos que describen recursos en la Web.
- WSML (Web Service Modeling Lenguaje): es el lenguaje para el modelado de Servicios Web Semánticos basado en WSMO (Web Service Modeling Ontology), el lenguaje WSML consta de una serie de lenguajes accesorios que definen determinados aspectos referentes a la semántica del servicio web, estos son: WSML-Core, WSML-DL, WSML-Flight, SWRL y WSML-Full

2.5.2. RDF: Resource Description Framework

RDF son las siglas utilizadas para referirse al término anglosajón *Resource Description Framework*. Es un modelo de metadatos para la publicación, la vinculación y el intercambio de datos y la descripción de recursos en la Web. Surgió en agosto de 1997 desde el Consorcio Web W3C y se convirtió en una recomendación de dicho consorcio en 2004 (Corlosquet, 2009). En él confluyen lenguaje de marcas y metadatos, proporcionando una arquitectura genérica de metainformación expresada en XML, que se implementará en todos los componentes (navegadores, servidores, etc.) que conforman la infraestructura web a fin de proporcionar un marco estándar para la interoperabilidad en la descripción de contenidos web sin necesidad de intervención humana (Méndez Rodríguez, 1999, Senso Ruiz, 2003).

Por dichas razones, Yu (2011) se refiere a él como el bloque de construcción de la Web Semántica, pues permite estructurar la información en forma de tripletas (Yu, 2011).

A diferencia de HTML (que se adapta para la visualización de información y hacerla entendible a los humanos), RDF está diseñado para ser leído y entendido por las máquinas, lo que permite el intercambio de datos. Se puede utilizar en una variedad de áreas dentro del procesamiento de la información (desde la visión de la CI); por ejemplo: en la búsqueda de recursos para proporcionar mejores capacidades de los motores de búsqueda, en la catalogación para describir el contenido y las relaciones de contenido disponible en un sitio en particular Internet, página, o biblioteca digital, por parte de agentes de software inteligentes para facilitar el intercambio de conocimientos y el intercambio, en la clasificación de contenidos, en la descripción de colecciones de páginas que representan una única lógica "documento"; para describir los derechos de propiedad intelectual de las páginas Web; para expresar las preferencias de privacidad de un usuario, así como las políticas de privacidad de un sitio Web. RDF con firmas digitales será clave para la construcción de la "Web de confianza" en el comercio electrónico, la colaboración y otras aplicaciones (W3C, 1999). Es la infraestructura que permite la restricción, reutilización e intercambio de metadatos estructurados.

Tiene características que facilitan la fusión de datos, incluso si los esquemas subyacentes difieren, y soporta específicamente la evolución de esquemas en el tiempo sin necesidad de todos los consumidores de datos que pueden cambiar.

RDF está estrechamente relacionado con las redes semánticas (Heflin, 2007). Se pueden destacar tres aspectos de su semántica funcional: “un modelo de datos, una sintaxis y un esquema” (Méndez Rodríguez, 1999), pues RDF define un modelo basado en tripletas: (ver fig.2.12)

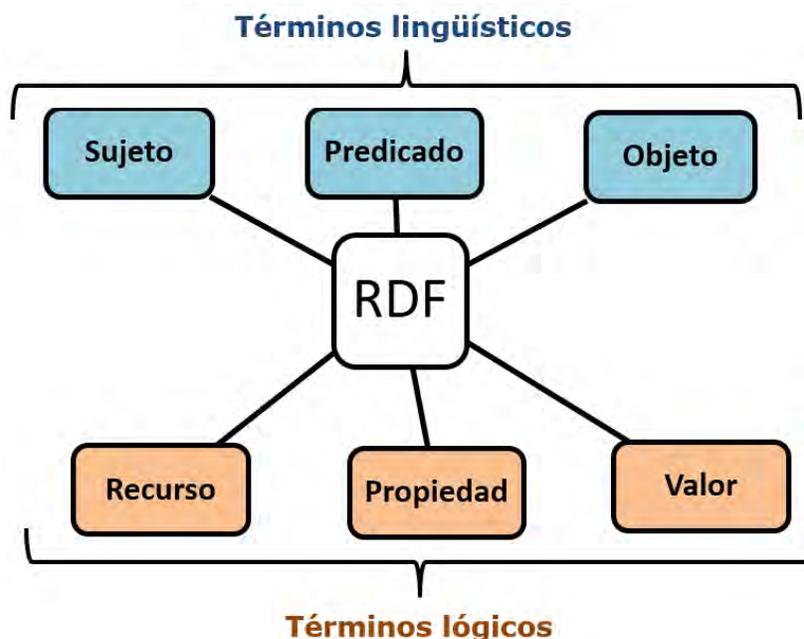


Fig. 2.12. Componentes de RDF: equivalencias lógico-lingüísticas²⁹ (elaboración propia)

²⁹ Basado en:

- PEDRAZA JIMÉNEZ, R., CODINA BONILLA, L. & ROVIRA, C. 2007. Web semántica y ontologías en el procesamiento de la información documental. *El profesional de la información*, 16, 569-578.
- MÉNDEZ RODRÍGUEZ, E. M. Year. RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio. *In: Les biblioteques i els centres de documentació al segle XXI: peça clau de la societat de la informació*, 1999. 487-498.
- W3C. 1999. *Resource Description Framework (RDF): Modelo y Sintaxis Especificación* [Online]. Available: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222> [Accessed 2015].

Como se muestra en la fig.2.12, RDF asume: (Méndez Rodríguez, 1999, W3C, 1999, Pedraza-Jimenez et al., 2007, Heflin, 2007);

- Recursos: Todas las cosas están descritas por expresiones RDF se denominan recursos, es decir, cualquier recurso entendido como objeto de información en la web, identificable unívocamente por un URI (identificador uniforme de recursos): documento, una parte o XML dentro de un documento fuente, colección de páginas, sitio web completo, etc. También pueden ser asumidas como tal: personas, o cualquier objeto del mundo real o conceptual.
- Propiedades: son aspectos específicos, características, atributos o relaciones utilizadas para describir **recursos**. Cada tipo de propiedad tiene sus valores específicos, define los valores permitidos, los tipos de recursos que puede describir y las relaciones que existen entre las distintas propiedades. Determina qué tipo de relación se mantiene entre el sujeto y el objeto.
- Valores: son los datos en los que se concreta la propiedad de un recurso determinado. Puede ser alternativamente tanto un valor literal como una cadena o un número.

Con ellos se construyen declaraciones, sentencias o descripciones que no son más que el conjunto de un recurso, un nombre de propiedad y el valor de esa propiedad —sujeto, predicado y objeto, respectivamente. Las relaciones establecidas en la tripleta RDF pueden ejemplificarse de la siguiente forma: (ver fig. 2.13)

- HEFLIN, J. 2007. An Introduction to the OWL Web Ontology Language. Pensilvania: Lehigh University. National Science Foundation (NSF).

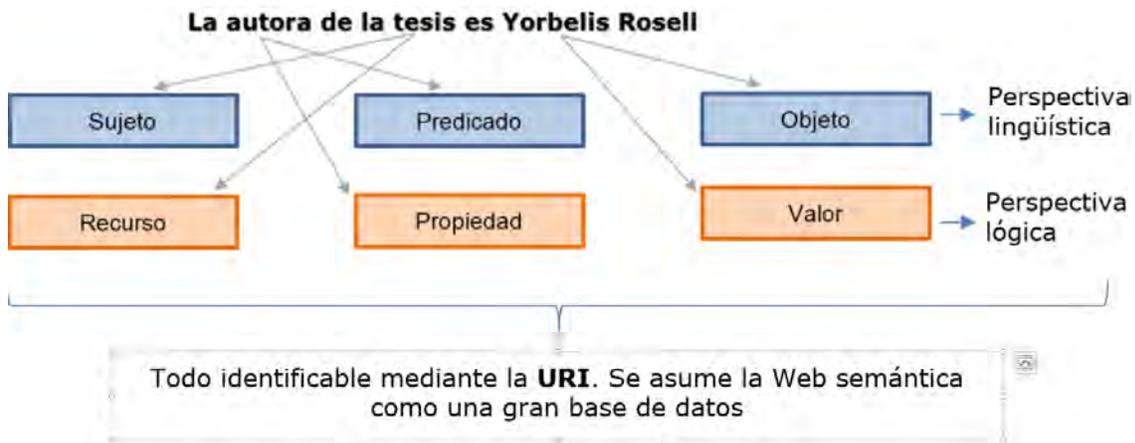


Fig. 2.13. Tripletas RDF: ejemplificación de las equivalencias lógico-lingüísticas (elaboración propia)

Por otra parte, RDF permite definir agrupaciones y clases con diferentes niveles de granularidad (Ver tabla 2.1.) y con RDF Schema³⁰ se proporciona un vocabulario que permite definir las clases para describir recursos y definir las propiedades que los recursos puedan tener.

Tabla 2.1. Formas de agrupar y clasificar información en RDF (elaboración propia)

<p>Clases y propiedades</p>	<p>Ej. Camila pertenece al grupo de investigación de Especies Marinas que da clases en la UH</p> <p><i>Camila</i> es una "persona"</p> <p>"profesores" y "grupoDeInvestigación" son clases</p> <p>"daClaseDe" y "perteneceA" son propiedades</p>
<p>Jerarquías de clases y herencia</p>	<p>Ej. La clase "profesores" está incluida en "personas"</p>
<p>Jerarquías de propiedades</p>	<p>Ej. "esProfesora" es una subpropiedad de "esTrabajadora"</p>

³⁰RDF Schema, "incluyerd`dfs:Class`, `RDF:Property` (del RDF namespace), `rddfs:subClassOf`, `rddfs:subPropertyOf`, `rddfs:domain`, y `rddfs:range`. También propiedades para la documentación, incluyendo `rddfs:label` y `rddfs:comment`." HEFLIN, J. 2007. An Introduction to the OWL Web Ontology Language. Pensilvania: Lehigh University. NationalScienceFoundation (NSF).

RDF aunque es solamente un modelo de datos y por sí mismo no tiene implicaciones semánticas (Heflin, 2007), se vislumbra, por su solidez y grado de implantación, como el instrumento para representar vastas cantidades de datos. Éstos pueden ser utilizados directamente por aplicaciones, sin que tengan que realizarse costosos procesos de conversión o traducción, puesto que se utilizan mecanismos compartidos de representación (Tramullas, 2011).

Otras estructuras o declaraciones complementarias que ofrece RDF son las siguientes (Pastor-Sánchez et al., 2012):

- Contenedores: Permiten la descripción de grupos de elementos. Pueden describir grupos en donde el orden de los elementos no es importante (Bag), otros contenedores en donde el orden de los elementos es relevante (Sequence) y grupos con elementos alternativos de los cuales puede seleccionarse uno para asignarlo a una propiedad (Alternative).
- Colecciones: Permiten crear grupos cerrados de elementos. Esta estructura al contrario que un Contenedor, que es abierto, permite enumerar de un modo preciso los elementos que la componen.
- Modelización: Permiten describir con RDF declaraciones de alto nivel con la intención de expresar conocimiento sobre otras declaraciones. En este caso la declaración se contempla como un recurso, representando explícitamente al sujeto, predicado, objeto y tipo de la afirmación.

A partir de los elementos vistos, en la fig.2.14 se presentan los constructores RDF y RDF Schema que puede agruparse en categorías funcionales.

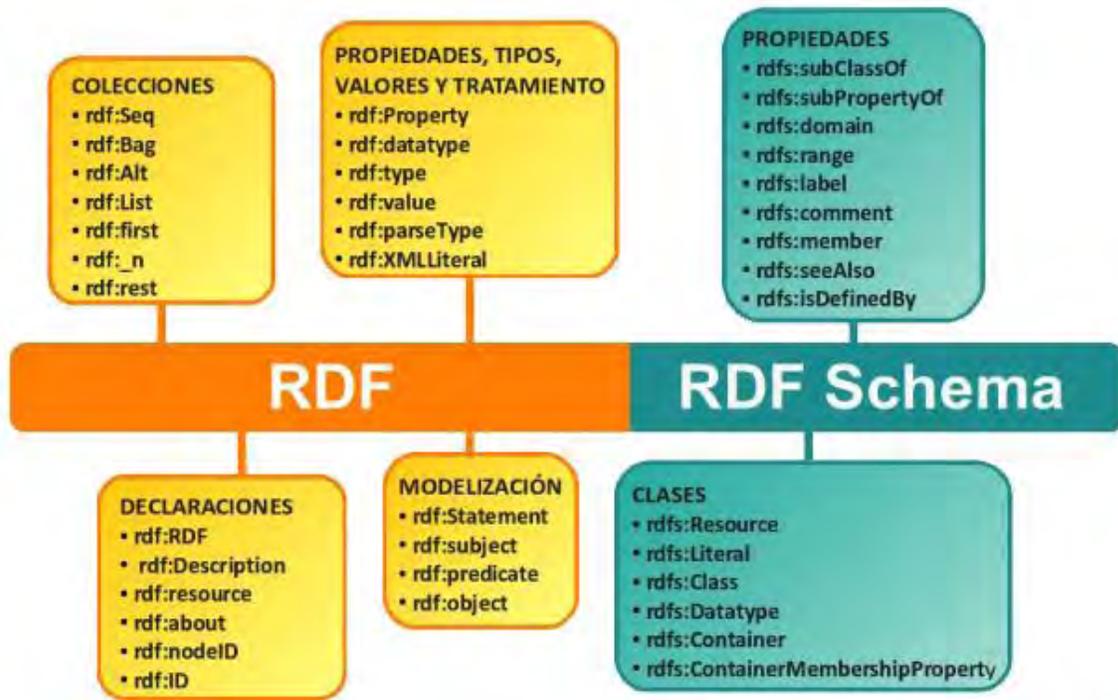


Fig. 2.14. Constructores RDF y RDF Schema (Pastor-Sánchez et al., 2012)

Por otra parte, el grupo W3C, ha publicado en su página, la relación de documentos normativos para el trabajo con RDF:

- Plataforma de datos vinculada 1.0 (Speicher et al., 2014): Esta especificación describe el uso de HTTP para acceder, actualizar, crear y eliminar recursos de los servidores que exponen sus recursos como Linked Data. Proporciona aclaraciones y ampliaciones de las reglas de Linked Data. Describen su estado utilizando RDF.
- Esquema RDF 1.1 (Brickley and Guha, 2014): El documento describe cómo utilizar RDF para describir vocabularios RDF. Esta especificación define un vocabulario para este propósito y define otros elementos de construcción en RDF vocabulario especificado inicialmente en el modelo RDF y especificación de sintaxis.
- RDF 1.1 XML Syntax (Gandon and Schreiber, 2014): Este documento define una sintaxis XML para RDF llamado RDF / XML en términos de

espacios de nombres en XML, el XML y el conjunto de información XML Base. La gramática formal de la sintaxis se anota con acciones de generación de tripletas de la gráfica RDF como se define en **RDF Conceptos y Sintaxis Abstracta** (Klyne et al., 2014). Las tripletas se escriben utilizando el formato gráfico de serialización N-Triples RDF que permite la grabación más precisa de la asignación en una forma procesable máquina. Las asignaciones se registran como pruebas de los casos, recogidos y publicados en RDF casos de prueba.

- RDF 1.1 Conceptos y Sintaxis Abstracta (Klyne et al., 2014): El documento define una sintaxis abstracta (un modelo de datos) que sirve para enlazar todos los idiomas y especificaciones basadas en RDF. La sintaxis abstracta tiene dos estructuras de datos clave: grafos RDF son conjuntos de triples sujeto-predicado-objeto, en donde los elementos pueden ser nodos iris, espacio en blanco, o literales datatyped. Se utilizan para expresar descripciones de recursos. RDF conjuntos de datos se utilizan para organizar colecciones de grafos RDF, y comprenden un gráfico predeterminado y cero o más nombrado gráficos. Este documento también presenta los conceptos clave y la terminología, y el manejo de los identificadores de fragmentos en los IRI dentro de grafos RDF.
- RDF 1.1 Turtle (Beckett et al., 2014): El Marco de Descripción de Recursos (RDF) es un lenguaje de propósito general para representar la información en la Web.
- RDF 1.1 N-Quads (Carothers, 2014): N-Quads es un formato de texto plano para codificar un conjunto de datos RDF. Similar a RDF 1.1 N-Triples, pero la principal distinción es que las N-Quads permiten codificar múltiples gráficos.
- RDF 1.1 N-Triples(Beckett, 2014): N-Triples es un formato de texto plano para codificar un grafo RDF. Se definió originalmente como una sintaxis para los Casos de Prueba RDF.

- RDF 1.1 Semantics (Hayes and Patel-Schneider, 2014): Este documento describe una semántica precisa para la Descripción de Recursos Framework 1.1 y Esquema RDF. En él se definen una serie de regímenes de vinculación distintas y correspondientes patrones de vinculación. Es parte de un conjunto de documentos que componen la especificación completa de RDF 1.1.
- RDF 1.1 TriG (Bizer and Cyganiak, 2014): Este documento define una sintaxis textual para RDF llamado TriG que permite a un conjunto de datos RDF para ser completamente escrito en una forma compacta de texto y natural, con las abreviaturas de los patrones de uso comunes y tipos de datos. TriG es una extensión del formato RDF 1.1 Turtle (Beckett et al., 2014).
- JSON-LD 1.0 Algoritmos de Procesamiento y API (Longley et al., 2015): Una interfaz de programación de aplicaciones y un conjunto de algoritmos para la transformación mediante programación documentos JSON-LD con el fin de hacerlos más fáciles de trabajar en entornos de programación como JavaScript, Python y Ruby.
- JSON-LD 1.0 (Sporny et al., 2014): Un formato común representación JSON para expresar gráficos dirigidos. Mezcla de Linked Data y non-Linked Data en un solo documento JSON.
- rdf:PlainLiteral: A Datatype for RDF Plain Literals (Segunda Edición)(Bao and Motik, 2009): En este documento se presenta la especificación de un tipo de datos primitivo de los literales simples de RDF.
- RDF Semantics (Hayes and McBride, 2004): Se trata de una especificación de una semántica precisa, y sistemas completos de reglas de inferencia, para la Descripción de Recursos (RDF) y el Esquema RDF (RDFS) correspondiente.
- RDF casos de prueba(Grant and Beckett, 2004): Este documento describe los casos de prueba RDF a entregar por el Grupo de Trabajo Core RDF como se define en la Carta del Grupo de Trabajo.

- Resource Description Framework (RDF): Concepts and Abstract Syntax (Klyne and Carroll, 2004): Define una sintaxis abstracta en la que se basa RDF, y que sirve para enlazar su sintaxis concreta a su semántica formal. También incluye la discusión de los objetivos de diseño, conceptos clave, datatyping, normalización de caracteres y manejo de referencias URI.
- RDF Primer (Manola et al., 2004): Este manual está diseñado para proporcionar al lector con los conocimientos básicos necesarios para utilizar con eficacia RDF. Se introduce los conceptos básicos de RDF y describe su sintaxis XML. En él se describe cómo definir vocabularios RDF utilizando el vocabulario de lenguaje de descripción, y da una visión general de algunas aplicaciones RDF desplegadas. También se describe el contenido y la finalidad de otros documentos de especificación RDF.

Sin embargo, RDF y RDF Schema no son suficientes para el manejo de la representación de principios semánticos en la Web, lo que constituyó la causa fundamental para el desarrollo de otros estándares como OWL o SKOS que se basan en RDF y proporcionan lenguaje para definir ontologías estructuradas, que permiten la integración y la interoperabilidad de los datos entre las comunidades descriptivas.

2.5.3. OWL: Web Ontology Language

OWL son las siglas utilizadas para referirse al término anglosajón *Web Ontology Language*. Es un estándar internacional para la codificación e intercambio de ontologías, basado en lógicas descriptivas que favorecen el desarrollo de aplicaciones semánticas.

Surge como propuesta de la W3C desde febrero de 2004 para apoyar la en desarrollo del proyecto de la Web semántica. Por tanto, como otros estándares de este grupo, desde su génesis es diseñado para ser compatible con XML. Sintácticamente, una ontología OWL es un documento RDF y, consecuentemente con ello, un documento XML bien estructurado. Esto permite que OWL sea procesado por la amplia gama de herramientas XML y RDF disponibles. "Cada uno de los términos RDF Schema importantes están

incluidos en OWL o está sustituido por nuevos términos” (Heflin, 2007). **Además, “OWL asigna un significado adicional a ciertos triples RDF”** (Dean et al., 2004), aumentando la expresividad de este estándar.

Desde un punto de vista técnico, OWL está diseñado para ser utilizado por aplicaciones que necesitan para procesar el contenido de la información en lugar de sólo presentar la información a los seres humanos. Facilita la interpretación del contenido Web a la máquina, proporcionando el vocabulario adicional junto con una semántica formal (Dean et al., 2004). OWL tiene tres sublenguajes expresivos: (Ver fig. 2.15)

- OWL Lite: Es una extensión de RDFs e incorpora los elementos más comunes de OWL, está destinado a usuarios que solamente necesitan crear taxonomías de clases y restricciones simples.
- OWL DL: Incluye completamente el vocabulario OWL. Provee 40 elementos primitivos (16 clases y 24 propiedades). OWL DL impone una serie de restricciones sobre grafos RDF. Asegurar que un documento RDF / XML cumple con todas las restricciones es una tarea relativamente difícil para los autores, y muchos documentos existentes son nominalmente OWL (Parsia and Sirin, 2004).
- OWL Full: Provee mayor flexibilidad para representar ontologías que OWL DL. Es esencialmente una ampliación de RDF con vocabulario adicional, sin restricciones en la forma en que se utiliza ese vocabulario (Bechhofer, 2009).

Property name	domain	range
owl:intersectionOf	owl:Class	rdf:List
owl:unionOf	owl:Class	rdf:List
owl:complementOf	owl:Class	owl:Class
owl:oneOf	owl:Class	rdf:List
owl:onProperty	owl:Restriction	rdf:Property
owl:allValuesFrom	owl:Restriction	rdfs:Class
owl:hasValue	owl:Restriction	<i>not specified</i>
owl:someValuesFrom	owl:Restriction	rdfs:Class
owl:minCardinality	owl:Restriction	xsd:nonNegativeInteger OWL Lite: {0,1} OWL DL/Full: {0,...,N}
owl:maxCardinality	owl:Restriction	xsd:nonNegativeInteger OWL Lite: {0,1} OWL DL/Full: {0,...,N}
owl:cardinality	owl:Restriction	xsd:nonNegativeInteger OWL Lite: {0,1} OWL DL/Full: {0,...,N}
owl:inverseOf	owl:ObjectProperty	owl:ObjectProperty
owl:sameAs	owl:Thing	owl:Thing
owl:equivalentClass	owl:Class	owl:Class
owl:equivalentProperty	rdf:Property	rdf:Property
owl:sameIndividualAs	owl:Thing	owl:Thing
owl:differentFrom	owl:Thing	owl:Thing
owl:disjointWith	owl:Class	owl:Class
owl:distinctMembers	owl:AllDifferent	rdf:List
owl:versionInfo	<i>not specified</i>	<i>not specified</i>
owl:priorVersion	owl:Ontology	owl:Ontology
owl:incompatibleWith	owl:Ontology	owl:Ontology
owl:backwardCompatibleWith	owl:Ontology	owl:Ontology
owl:imports	owl:Ontology	owl:Ontology

Fig. 2.15. Propiedades de OWL (Gómez-Pérez et al., 2006)

Los diferentes constructores de OWL Lite, OWL DL y OWL Full puede agruparse en una serie de categorías funcionales tal y como se muestra en la siguiente ilustración. (Ver fig. 2.16)

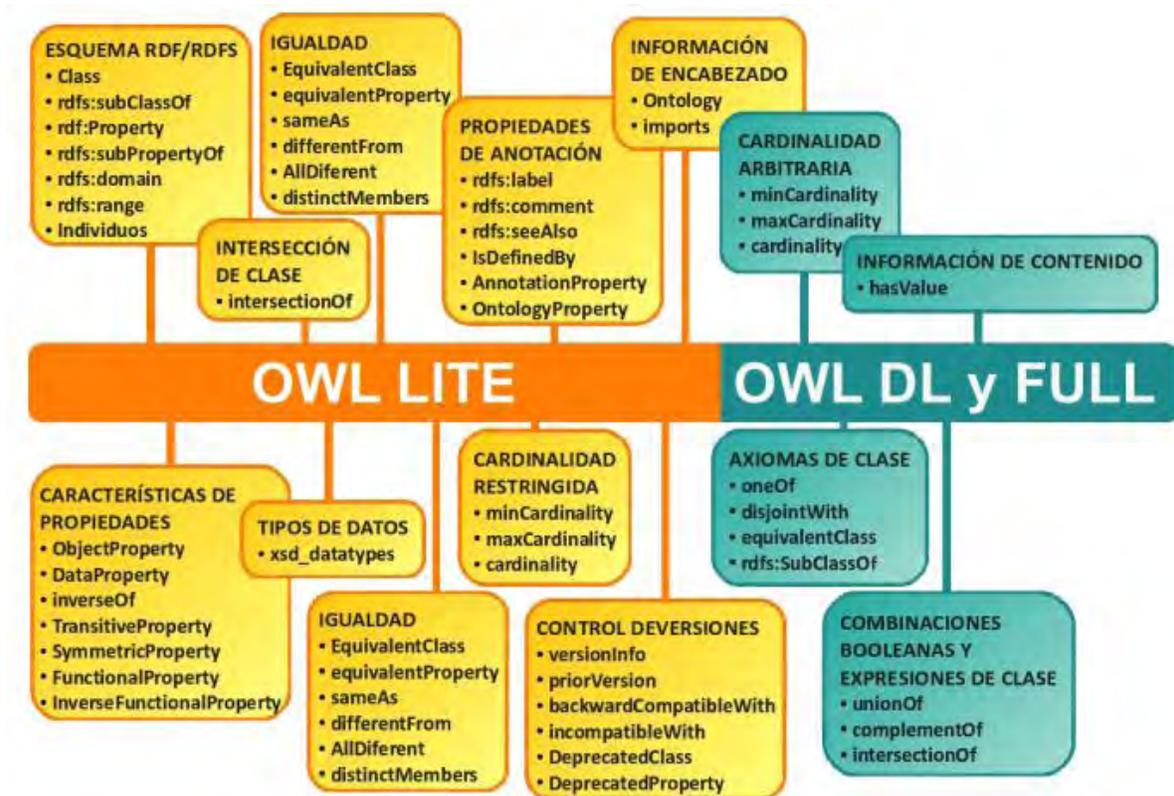


Fig. 2.16. Constructores OWL Lite, DL y Full (Pastor-Sánchez et al., 2012)

Podría decirse que OWL Full se contempla como una extensión de RDF mientras que OWL Lite y OWL DL pueden considerarse extensiones de una visión restringida de RDF. Es decir, todo documento de OWL es un documento RDF y cada documento RDF es un documento OWL Full, pero sólo algunos documentos RDF pueden considerarse documentos OWL Lite o DL (Pastor-Sánchez et al., 2012).

OWL provee axiomas que dotan de restricciones a las clases definidas y las relaciones permitidas entre estos elementos. Como lenguaje para la construcción de ontologías permite realizar especificaciones a niveles abstractos (conceptual) del conocimiento de un dominio. Además utiliza la lógica descriptiva. Entre las principales bondades que exhibe pueden mencionarse (Ruiz, 2008):

- Restricciones de rango (las vacas sólo comen hierba)

- Clases disjuntas (ej. hombres y mujeres)
- Combinaciones booleanas (unión, intersección, etc.)
- Restricciones de Cardinalidad (una persona tiene 2 progenitores)
- Características de propiedades (transitividad, unicidad, inversa,)
- Existen técnicas de prueba para subconjuntos de lógica de primer orden
- Propiedades: Consistencia, Completitud, tratabilidad
- RDF(S) y OWL son subconjuntos de lógica descriptiva
- Otros formalismos: Cláusulas
- Lenguajes de reglas en desarrollo.

Por dichas razones, a decir de algunos autores, OWL, es el lenguaje de mayor significado semántico, así como el que mayor libertad otorga para implementar una ontología (Castro Fernández, 2008).

El lenguaje OWL se describe a lo largo de varios documentos publicados en la página de W3C:

- RIF RDF y OWL de compatibilidad (Segunda Edición)(McGuinness and Van Harmelen, 2004): Una especificación formal de cómo RIF se puede utilizar con RDF y OWL, incluyendo la semántica de las diferentes formas de importar datos RDF y ontologías OWL en los sistemas de reglas de RIF.
- Ontologías Web OWL 2: Mapeo de idioma para RDF gráficos (Segunda Edición)(Grau et al., 2008): Este documento define el mapeo de ontologías OWL 2 en grafos RDF, y viceversa.
- OWL 2 Web Ontology Language New Features and Rationale (Segunda Edición)(Golbreich et al., 2009): Este documento es una simple introducción a las nuevas características de la Web OWL 2 lengua de la ontología, incluyendo una explicación de las diferencias entre la versión inicial de OWL y OWL 2. El documento también presenta los requisitos que han motivado

el diseño de las principales novedades, y su justificación desde una perspectiva teórica y aplicación.

- Ontologías Web OWL 2 perfiles lingüísticos (Segunda Edición)(Motik et al., 2009a): Este documento proporciona una especificación de varios perfiles de OWL 2 que pueden ser implementadas más sencilla y / o eficiente. En lógica, los perfiles son a menudo llamados fragmentos. La mayoría de los perfiles se definen por la imposición de restricciones sobre la estructura de OWL 2 ontologías. Estas restricciones se han especificado mediante la modificación de las producciones de la sintaxis de estilo funcional.
- OWL 2 Web Ontology Language Quick Reference Guide (Segunda Edición) (Bao et al., 2009): Este documento proporciona una guía de referencia rápida al lenguaje OWL 2.
- OWL 2 Web Ontology Language RDF-Based Semantics (Segunda Edición)(Schneider et al., 2009): Este documento define la semántica de teoría de modelos compatibles con RDF de OWL 2.
- OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Segunda Edición)(Motik et al., 2009d): Las construcciones significativas proporcionadas por OWL 2 se definen en términos de su estructura. A su vez, una sintaxis de estilo funcional se define para estas construcciones, con ejemplos y descripciones informales. Se puede razonar con OWL 2 ontologies ya sea en virtud de la semántica basada en RDF o la semántica directa.
- OWL 2 Web Ontology Language XML Serialization (Segunda Edición)(Motik et al., 2009b): Este documento especifica una serialización XML para OWL 2 que refleja su especificación estructural. Un esquema XML define la sintaxis y está disponible como un documento separado, además de ser incluido aquí.
- OWL 2 Web Ontology Language Conformance (Segunda Edición)(Horrocks et al., 2009): Este documento describe las condiciones que OWL 2 herramientas deben cumplir para ser conformes con la especificación del

lenguaje. También presenta un formato común para los casos de prueba OWL 2 ilustran que tanto las características de la lengua y puede ser utilizado para las pruebas de conformidad.

- OWL 2 Web Ontology Language Direct Semantics (Segunda Edición)(Motik et al., 2009c): Este documento proporciona la semántica directos de teoría de modelos de OWL 2, que es compatible con el SROIQ descripción de la lógica. Por otra parte, este documento define los problemas de inferencia más comunes para OWL 2.
- OWL 2 Web Ontology Language Document Overview (Segunda Edición)(W3C, 2012a): Este documento sirve como una introducción a OWL 2 y de los demás documentos OWL 2. En él se describe la sintaxis para OWL 2, los diferentes tipos de semántica, los perfiles disponibles (sub-idiomas), y la relación entre OWL 1 y 2.
- OWL Web Ontology Language Overview(McGuinness and Van Harmelen, 2004): ÓWL está diseñado para ser utilizado por las aplicaciones que necesitan para procesar el contenido de la información en lugar de simplemente presentar la información a los seres humanos. OWL facilita una mayor interpretabilidad para las máquinas sobre el contenido Web que se soportado en XML, RDF y RDF Schema (RDF-S), proporcionando el vocabulario adicional junto con una semántica formal. OWL tiene tres sublenguas cada vez más expresivos: OWL Lite, OWL DL y OWL Full. Este documento está escrito para los lectores que quieran una primera impresión de las capacidades de OWL. Proporciona una introducción a OWL, describiendo de manera informal las características de cada uno de los sublenguajes de OWL. Algunos de los conocimientos de RDF-S son útiles para la comprensión de este documento, pero no es esencial.

La capacidad de la Web Semántica para vincular información de múltiples fuentes es una característica deseable y potente que se puede utilizar en muchas aplicaciones. Sin embargo, la capacidad para combinar datos de múltiples fuentes, combinadas con el poder inferencial de OWL, tiene potencial para el abuso. Incluso puede ser ilegal para crear o procesar dicha información

vinculada en países con leyes de protección de datos (...). Por lo tanto, debe tenerse especial cuidado al usar OWL con cualquier tipo de datos de carácter personal que pudiera estar relacionado con otras fuentes de datos u ontologías (Dean et al., 2004).

2.6. Ontologías y Web semántica

Desde principios de la década de los 90, las ontologías se han convertido en un tema de investigación popular. Los investigadores del campo de la inteligencia artificial fueron los primeros en desarrollar ontologías, y hoy varias comunidades de investigación, incluyendo la ingeniería del conocimiento, la adquisición de conocimientos, procesamiento de lenguaje natural y representación del conocimiento, entre otros, trabajan sobre ellas. En dicho sentido, diferentes áreas de estudio la han abordado: la ingeniería de software, las matemáticas, la informática y más recientemente en el campo de las Ciencias de la Información, como herramientas en la representación de información (como esquema conceptual), en búsqueda y recuperación (como herramienta), como sistemas de información cooperativos, y su aplicación a bibliotecas digitales, comercio electrónico y gestión del conocimiento.

Ejemplo de esto, lo constituyen los diversos proyectos desarrollados en Internet con lenguajes de codificación de ontologías (Rodríguez Perojo and Ronda León, 2005):

- El servidor Ontolingua, resultado del KSE (Knowledge Sharing Effort), ofrece herramientas para crear ontologías, integrarlas con otras existentes e incorporarlas a nuevos productos de software.
- El desarrollo de SHOE (Simple HTML Ontology Extensions), aportado por Luke, Spector y Rager con como complemento semántico de HTML, que refleja el contenido de la página web y que puede utilizarse por agentes de software para el descubrimiento de información. Posteriormente SHOE ha evolucionado hacia RDF, y es OWL (Web Ontology Language), la más reciente especificación, mantenida por el Consorcio del Web y que intenta proveer de un lenguaje que pueda utilizarse para describir

clases y relaciones entre ellas inherentes a documentos y aplicaciones web.

- El proyecto FERMI (Formalization and Experimentation on the Retrieval Multimedia Information), bajo la supervisión de C. J "Keith" Van Rijsbergen, en el que se incluyen herramientas de planificación, descubrimiento y selección de recursos de información multimedia.
- El proyecto IMP (Information Manifold Project) desarrollado en el ámbito de Bell Labs, en el que se hace uso de las ontologías para identificar las fuentes de información pertinentes a una búsqueda, acceder a ellas, obtener documentos relevantes, compararlos, seleccionar los más adecuados y ofrecer un resumen previo al usuario.
- El proyecto UMLS (Unified Medical Language System), en el ámbito de la medicina, desarrollado por la National Library of Medicine de los Estados Unidos, que utiliza las ontologías como una herramienta más para el acceso, integración y recuperación de información biomédica.

Las ontologías han sido consideradas la columna vertebral de la Web Semántica. Como ya se había abordado en el epígrafe referido a la estructura de la Web semántica, **la capa de la ontología "describe e identifica lo que nuestros conceptos significan para que podamos utilizar los metadatos para crear servicios de integración, federar, compartir y procesar los datos. La ontología da un entendimiento compartido y que permite la reutilización de conocimiento del dominio"** (Antoniou et al., 2005).

La integración de las ontologías a los CMS, dieron al traste con los CMS semánticos desde donde se pueden definir vocabularios de términos como parte de la Web semántica con el uso de RDF y OWL (W3C, 2012b).

Aunque no todo el mundo tiene que compartirlas, para las organizaciones una ontología es útil para visualizar y comunicar el dominio. Esto hace que sea más fácil para los administradores, editores, diseñadores, etc., de un sistema de información implementado en la Web desde el CMS, asegurar que la estructura de la página sea estable, y que permita modificar el modelo sin cambiar la

base de datos o volver a escribir ningún código, con total independencia en el manejo del diseño, el contenido y la semántica.

2.6.1. Ontología: conceptualización del término

El término ontología, aunque proviene de la filosofía, ha tenido un carácter marcadamente interdisciplinar con aportaciones desde la filosofía, la lingüística y las ciencias cognitivas en general, lo que ha derivado en el aporte de definiciones desde diferentes áreas de conocimiento.

En el campo de la inteligencia artificial se identifican diferentes autores que tienen varias definiciones declarativas y procedimentales de ontologías.

La definición declarativa más consolidada, a decir de varios autores (Guerrero Bote and Lozano-Tello, 1999, Lozano-Tello et al., 2003, Giaretta and Guarino, 1995, Studer et al., 1998, García-Marco, 2007), es la propuesta por Gruber (Gruber, 1993a, Gruber, 1993b) que la describe como una especificación explícita y formal sobre una conceptualización compartida.

Neches (1991) define la ontología desde una visión procedimental al conceptualizarla como el vocabulario de un área mediante un conjunto de términos básicos y relaciones entre dichos términos, así como las reglas que combinan términos y relaciones que amplían las definiciones aportadas en el vocabulario. En una ontología además de las definiciones explícitas, existirán definiciones implícitas que podrán ser inferidas por el uso de axiomas (Neches et al., 1991).

Las ontologías vistas, específicamente, desde la rama relacionada con la representación del conocimiento, la ingeniería del conocimiento, que se ocupa de la construcción de sistemas expertos, a decir de García-Marco es un sistema de términos que sirve para describir y representar un área de conocimiento y que expresa las relaciones entre ellos por medio de un lenguaje formal que puede ser entendido por un ordenador (García-Marco, 2007).

A los efectos de la presente investigación se asume la ontología como una descripción de los conceptos y relaciones en un dominio de aplicación, descrito en un lenguaje equipado con una semántica formal compartida y consensuada;

legible y utilizable por los ordenadores. De esta forma se ofrecería un resumen o visión simplificada del mundo que queremos representar.

Por otra parte, Noy y McGuinness, señalan que, a partir de la conceptualización del término, las ontologías tendrían por objetivo (Noy and McGuinness, 2001):

- Intervenir en la comprensión común de la estructura de la información entre personas o agentes de software, lo que debe revertirse de forma positiva y casi necesaria en la extracción y recuperación de información, en páginas web, de contenidos conectados temáticamente.
- Permitir la reutilización del conocimiento perteneciente a un dominio.
- Explicitar los supuestos de un dominio para la representación del conocimiento y la información más allá de consideraciones técnicas, operativas e informáticas.
- Separar el conocimiento del dominio del conocimiento operacional. Con esto se alude a que, en ocasiones, el conocimiento que se está representando se puede implicar en diferentes áreas al pertenecer más a un conocimiento relacionado con procesos.
- Analizar el conocimiento de un campo, en lo que se refiere al estudio de los términos y relaciones que lo configuran ya sea formalmente o no.

Para lograr sus objetivos, las ontologías requieren de varios componentes que permiten representar el conocimiento de algún dominio (Gruber, 1993b):

- Conceptos: son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.
- Relaciones: representan la interacción y enlace entre los conceptos del dominio. Suelen formar la taxonomía del dominio. Por ejemplo: subclase-de, parte-de, parte-exhaustiva-de, conectado-a, etc.
- Funciones: son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como categorizar-clase, asignar fecha, etc.

conceptos relacionados (como tesauros y redes semánticas) pero permiten definir relaciones semánticas complejas, reglas y axiomas, que no están presentes en el resto de sistemas de organización del conocimiento.

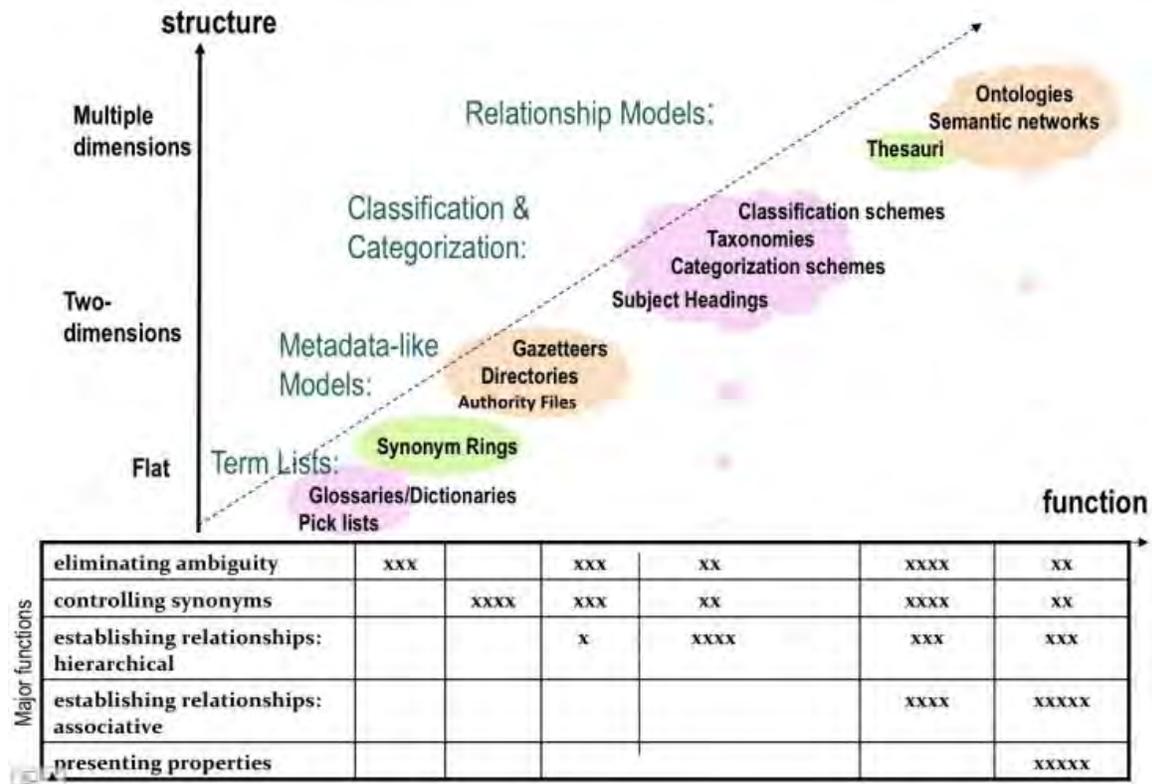


Fig. 2.18. Evolución de los diferentes tipos de sistemas organización del conocimiento (Zeng and Chan, 2004)

En ellas existe un nivel más alto de descripción del vocabulario y desarrollo semántico más profundo; relaciones semánticas complejas, reglas y axiomas que no están presentes en el resto de sistemas de organización del conocimiento. Además, favorecen el razonamiento inferencial de los agentes software. Suponen, hasta el momento, de los canales más eficientes para comunicación entre personas y máquinas que tributa a la recuperación de conocimiento con la ayuda de agentes software.

Varios autores (Guerrero Bote and Lozano-Tello, 1999, Zeng and Chan, 2004) han coincidido sobre estos aspectos al señalar que las ontologías proporcionan un método más rico de organizar el conocimiento, permitiendo realizar

deducciones sobre conocimiento formalizado que puede además ser reutilizado, lo que constituye un gran avance en el entendimiento de diversos sistemas de representación del conocimiento y llevarán a fomentar el desarrollo de aplicaciones basadas en dichos sistemas.

A pesar de ello, es importante advertir, que la simple construcción de una ontología formal, desde la visión informática, no resuelve, aborda sistemáticamente, ni agota la totalidad de los problemas ontológicos en documentación ni en otras disciplinas adyacentes, especialmente el de las categorías universales que son del interés de la ontología filosófica, la semántica y la organización del conocimiento; o las leyes de la estructuración del significado que estudia la semántica. Las ontologías aportan una formalización de las meta-relaciones ontológicas que permite las operaciones lógicas entre ellas, pero la especificación de los entes, las restricciones y las relaciones específicas queda como un problema dentro de cada dominio y aun a nivel universal (García-Marco, 2007).

2.6.2. Tipos de ontologías

Las ontologías, como casi todo en la ciencia, tienen varias maneras de clasificarse para su estudio. Se pueden establecer distintos tipos de ontologías atendiendo a diversos aspectos. Podemos destacar las siguientes clasificaciones, aunque existen otras muchas:

Según su dependencia y relación con una tarea específica desde un punto de vista, Giaretta y Guarino clasifican las ontologías en (Giaretta and Guarino, 1995):

- **Ontologías de Alto nivel o Genéricas** describen conceptos más generales.
- **Ontologías de Dominio** describen un vocabulario relacionado con un dominio genérico.
- **Ontologías de Tareas o Técnicas básicas** describen una tarea, actividad o artefacto, por ejemplo componentes, procesos o funciones.

- **Ontologías de Aplicación** describen conceptos que dependen tanto de un dominio específico como de una tarea específica, y generalmente son una **especialización de ambas**”.

Según Van Heijst las ontologías pueden clasificarse de acuerdo a la cantidad y tipo de estructura de la conceptualización en: (Van Heijst et al., 1997, Guerrero Bote and Lozano-Tello, 1999).

- **Ontologías terminológicas** que especifican los términos que son usados para representar conocimiento en el universo del discurso. Suelen ser usadas para unificar vocabulario en un campo determinado.
- **Ontologías de información** especifican la estructura de almacenamiento de bases de datos. Ofrecen un marco para almacenamiento estandarizado de información.
- **Ontologías del modelado del conocimiento** especifican conceptualizaciones del conocimiento. Contiene una rica estructura interna y suelen estar ajustadas al uso particular del conocimiento que describen.

Según Lamarca pueden clasificarse atendiendo al ámbito del conocimiento al que se apliquen, el tipo de agente o usuario al que va dirigida y el nivel de abstracción y razonamiento lógico de las conexiones de su contenido.

Según el ámbito del conocimiento al que se apliquen (Lamarca, 2006):

- **Ontologías generales:** son las ontologías de nivel más alto ya que describen conceptos generales (espacio, tiempo, materia, objeto, etc.)
- **Ontologías de dominio:** describen el vocabulario de un dominio concreto del conocimiento.
- **Ontologías específicas:** son ontologías especializadas que describen los conceptos para un campo limitado del conocimiento o una aplicación concreta.

Según el tipo de agente al que vayan destinadas: (Lamarca, 2006)

- **Ontologías lingüísticas:** se vinculan a aspectos lingüísticos, esto es, a aspectos gramáticos, semánticos y sintácticos destinados a su utilización por los seres humanos.
- **Ontologías no lingüísticas:** destinadas a ser utilizadas por robots y agentes inteligentes.
- **Ontologías mixtas:** combinan las características de las anteriores.

Según el grado o nivel de abstracción y razonamiento lógico que permitan (Lamarca, 2006):

- **Ontologías descriptivas:** incluyen descripciones, taxonomías de conceptos, relaciones entre los conceptos y propiedades, pero no permiten inferencias lógicas.
- **Ontologías lógicas:** permiten inferencias lógicas mediante la utilización de una serie de componentes como la inclusión de axiomas, etc.

2.6.3. Metodologías para la construcción de ontologías

Desde el punto de vista metodológico, el proceso de desarrollo de una ontología, implican el conjunto de tareas que debería llevarse a cabo cuando éstas se construyen. Para la construcción de ontologías en los últimos 20 años se han desarrollado diversos principios y métodos que exponen cómo construirlas.

La más antigua es Cyc (Guzmán Luna et al., 2012, Carrillo Pozas, 2014) en la que divulgaron algunos pasos generales para la construcción de ontologías. El primero consiste en extraer manualmente el conocimiento común que está implícito en diferentes fuentes para después, cuando se tenga suficiente conocimiento en la ontología, adquirir nuevo conocimiento común usando herramientas de procesamiento de lenguaje natural o aprendizaje computacional (Guzmán Luna et al., 2012).

La metodología de Uschold y King se basa en la experiencia de construir el Enterprise Ontology, el cual incluye un conjunto de ontologías para modelado

empresarial. Ellos proponen los siguientes pasos: (1) Identificar el propósito y el alcance de la ontología; (2) Construir la ontología capturando conocimiento, codificando conocimiento e integrando tal conocimiento con ontologías existentes; (3) Evaluar la ontología; (4) Documentación; y (5) Lineamientos para cada fase (Uschold and King, 1995).

Paralelamente, surge la metodología de Grüninger y Fox usada para construir las ontologías del proyecto TOVE (Toronto Virtual Enterprise) (Grüninger and Fox, 1995). Constituye una aproximación lógica de primer orden para representar actividades, estados, tiempos, recursos y costes en una arquitectura de integración empresarial que incluye Enterprise Design Ontology, Project Ontology, Scheduling Ontology y Service Ontology. Esencialmente, plantea la elaboración de un modelo lógico del conocimiento que será especificado en la ontología. En este modelo primero, son descritas informalmente las especificaciones que tienen que ser encontradas por la ontología identificando un conjunto de asuntos competentes, y su descripción es entonces formalizada en un lenguaje basado en el cálculo de predicados de primer orden. Las preguntas de competencia son las bases para una caracterización rigurosa del conocimiento que tiene que cubrir la ontología, estas especifican el problema y lo que constituye una buena solución al problema. Con un mecanismo de composición y descomposición, las preguntas de competencia y sus respuestas pueden ser usadas para responder preguntas de competencia más complejas en otras ontologías, permitiendo la integración de ontologías. (Ver fig.2.19)

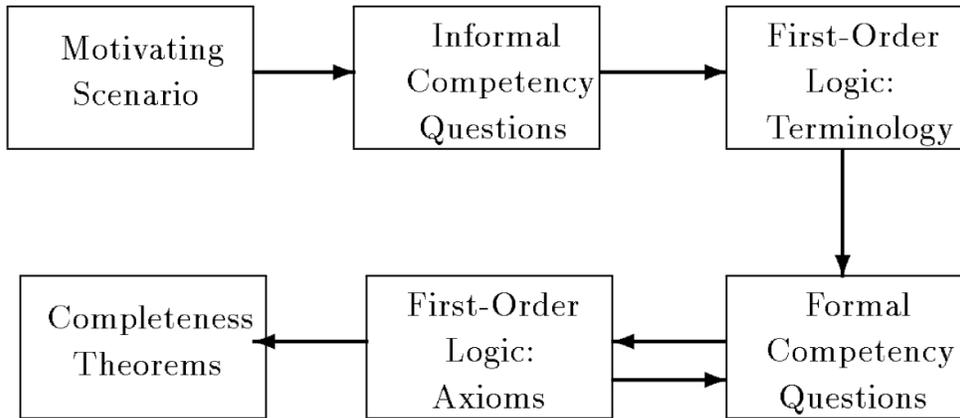


Fig. 2.19. Procedimiento para el diseño y evaluación de ontologías (Grüninger and Fox, 1995)

La metodología METHONTOLOGY es presentada en 1996 (Gómez-Pérez et al., 1996). **A decir de Guzmán Luna** "es una de las propuestas más completas ya que toma la creación de ontologías como un proyecto informático" (Guzmán Luna et al., 2012). Methodology permite la construcción de ontologías en el nivel del conocimiento y además permite construir ontologías totalmente nuevas o reutilizar otras ontologías. En sus etapas o fases incluye: la identificación del proceso de desarrollo de la ontología, un ciclo de vida propuesto y la metodología como tal. El proceso de desarrollo de la ontología identifica las tareas que deben realizarse cuando se construye una ontología (planeación, control, control de calidad, especificación, adquisición de conocimiento, conceptualización, integración, formalización, implementación, evaluación, mantenimiento, documentación y administración de la configuración). El ciclo de vida basado en prototipos evolutivos identifica las etapas que atraviesa la ontología durante su tiempo de vida. Finalmente, la metodología en sí misma, especifica los pasos que se deben seguir para desarrollar cada actividad, las técnicas usadas, los productos que se producirán y como serán estos evaluados. La fase principal en el proceso de desarrollo de una ontología usando el enfoque de METHONTOLOGY es la fase de conceptualización. Durante ambas: especificación y conceptualización, se completa un proceso de integración usando ontologías de la casa y ontologías externas. Finalmente, METHONTOLOGY propone que deben llevarse a cabo

actividades de evaluación a través de todo el tiempo de vida del proceso de desarrollo de la ontología.

Carrillo (2014) señala que a partir de METHONTOLOGY han sido desarrolladas varias ontologías usando esta metodología:

- CHEMICALS (López et al., 1999), una ontología en el dominio de los elementos químicos; Ontologías de contaminantes ambientales que representa los métodos de detección de los diferentes componentes contaminantes de varios medios: agua, aire, tierra, etc.; y la concentración máxima permitida de estos componentes, tomando en cuenta toda la legislación para este efecto (normas de la Unión Europea, normas Españolas, Alemanas, Estadounidenses, etc.).
- Ontology Reference (Arpirez et al., 1998) es una ontología en el dominio de las ontologías que juega el papel de una especie de páginas amarillas de ontologías
- Ontología KA (Blázquez et al., 1998) Esta metodología para construir ontologías ha sido propuesta por la FIPA (Foundation for Intelligent Physical Agents) que promueve la interoperabilidad entre aplicaciones basadas en agentes.

En ese mismo año (1996) un nuevo método fue propuesto para construir ontologías, este estaba basado en la ontología de SENSUS (Swartout et al., 1996), la cual constituye un enfoque top-down para derivar ontologías específicas del dominio a partir de grandes ontologías (ver fig.2.20).

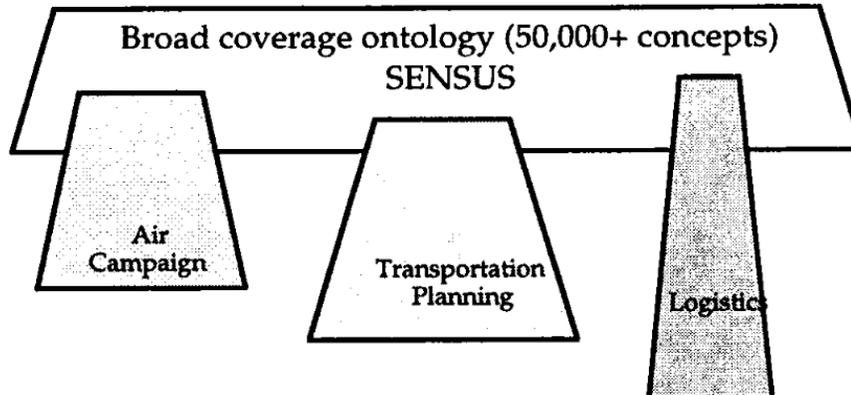


Fig. 2.20. Propuesta de SENSUS para vincular términos de dominio a una ontología de amplia cobertura (Swartout et al., 1996)

En esta metodología se identifican un conjunto de términos semilla que son relevantes en un dominio particular. Tales términos se enlazan manualmente a una ontología de amplia cobertura. Los usuarios seleccionan automáticamente los términos relevantes para describir el dominio y acotar la ontología Sensus. Consecuentemente, el algoritmo devuelve el conjunto de términos estructurados jerárquicamente para describir un dominio, que puede ser usado como esqueleto para la base de conocimiento.

Años más tarde es presentada "Methodologies for Ontology Development", de Jones, Bench-Caponand y Visser (1998), que a decir de Carrillo (2014), es de los primeros trabajos popularizados para el desarrollo de ontologías. En dicho documento los autores repasan las metodologías existentes e identifican los puntos en común que comparten las metodologías analizadas (Jones et al., 1998, Carrillo Pozas, 2014).

Corcho, Fernández, y Gómez-Pérez con "Methodologies, Tools and Languages for Building Ontologies: Where is the Meeting Point?", es el estudio al que desde la literatura especializada se le concede el segundo puesto en importancia. Tomó como base el análisis realizado en "Methodologies for Ontology Development" de Jones, Bench-Caponand y Visser (1998), al que añadió nuevos métodos. Incluyen cuadros comparativos detallados de

herramientas de desarrollo de ontologías, así como de lenguajes de ontologías. Proponen un diagrama de flujo de desarrollo de una ontología para: facilitar su desarrollo durante su ciclo de vida, su gestión y evolución y soporte (Carrillo Pozas, 2014, Corcho et al., 2003).

Existen otras propuestas, como "ONTOLOGY DEVELOPMENT 101" (cuya traducción del nombre es "Desarrollo de Ontologías-101) propuesta por la Universidad de Stanford (EEUU) (Noy and McGuinness, 2005). Esta metodología se distingue por la simplicidad, claridad y objetividad de los pasos establecidos. Consta de 7 pasos: (Ver fig. 2.21)



Fig. 2.21. Fases de la metodología ONTOLOGY DEVELOPMENT 101 (elaboración propia)

1. Determinar el dominio y alcance de la ontología

Sugiere responder a varias preguntas básicas, cómo: ¿Cuál es el dominio que la ontología cubrirá? ¿Para qué usaremos la ontología? ¿Para qué tipos de preguntas la in formación en la ontología deberá proveer respuestas? ¿Quién usará y mantendrá la ontología?

Deben preverse además el diseño de preguntas de competencia que puedan servir de medidor de la calidad de los resultados en la implementación de la ontología

2. Reutilización de ontologías existentes

Reusar ontologías existentes puede ser un requerimiento si el sistema necesita interactuar con otras aplicaciones que ya se han de dicado a

ontologías particulares o vocabularios controlados. Muchas ontologías ya están disponibles en forma electrónica y pueden ser importadas dentro un entorno de desarrollo de ontologías utilizadas.

3. Enumerar términos importantes

Resalta la utilidad de escribir una lista con todos los términos con los que se requiera hacer enunciados o explicaciones al usuario.

4. Definición de las clases y de la jerarquía de clases

No hay un orden único para definir las clases y establecer niveles de jerarquía. Puede partirse de lo general a lo específico (top-down) y viceversa (bottom-up).

5. Definir propiedades de las clases: SLOTS

Referida a la estructura interna de los conceptos definidos mediante clases y subclases en la etapa anterior.

Para cada propiedad en la lista, se debe determinar que clase es descrita por la propiedad. Esas propiedades se convierten en slots adosados a las clases.

6. Definir facetas de los slots

Los slots pueden tener diferentes facetas que describen el tipo de valor, valores admitidos, el número de los valores (cardinalidad), y otras características de los valores que los slots pueden tomar.

Bajo esta metodología se definen:

- **Cardinalidad de slot:** La cardinalidad de un slot **define** cuantos valores un slot puede tener.
- Tipo de valor: String, Number, Boolean (si son simples banderas si/no.); **del tipo Enumerated (donde se especifica una lista de valores admitidos para el slot);** y del **tipo Instance (que admiten la definición de relaciones entre individuos)**
- Dominio y rango:

- Dominio: Las clases a las cuales un slot está adosado o las clases cuyas propiedades son descritas por un slot son llamadas dominio del slot
- Rango: Las clases admitidas para los slots de tipo Instance son llamadas rango de un slot

7. Crear instancias

El último paso consiste en crear instancias individuales de clases en la **jerarquía**. La **definición de** una instancia individual de una clase requiere (1) elegir una clase, (2) crear una instancia individual de la clase y (3) rellenar los valores del slot.

En general, las metodologías para el desarrollo y construcción de ontologías son múltiples y muy diversas. La selección de una u otra, debe estar en consonancia con la arquitectura del sistema de información para el cual será construida la ontología, así como las características del dominio al cual responde.

2.6.4. Herramientas para la construcción de ontologías

En la elaboración y construcción de una ontología debe tener en cuenta su relación con la arquitectura del sistema de información en el que está inmersa sin olvidar la importancia que se deriva de la formular teorías de conocimiento sobre un dominio determinado.

Al igual que sucede con el abordaje de las metodologías, existen diversidad de herramientas destinadas a la construcción de las ontologías. Sin embargo, a decir de Corcho, las herramientas de construcción de ontologías son similares unas a otras, aun cuando muestren problemas de convergencia y de adaptación a los cambiantes lenguajes (Corcho et al., 2003).

Ejemplo de ellas son: Ontolingua Server, Ontosaurus encabezando la lista cronológica de las mismas, hasta las más recientes como DOGMA, KAON, OntoClean, OnToContent, HOZO, Protégé, WebODE, Ontoedit, entre otros.

Según criterios de algunos autores (García Jiménez, 2004, Gómez-Pérez et al., 2002) las herramientas para la elaboración de las ontologías se subdividen en:

1) Herramientas de desarrollo de ontologías: este grupo incluye las herramientas que sirven para la construcción de nuevas ontologías o bien para la reutilización de las existentes. Destacan entre sus funcionalidades la edición y la consulta, así como la exportación e importación de ontologías, la visualización en diversos formatos gráficos, etcétera.

2) Herramientas de fusión e integración de ontologías: pretenden solucionar el problema de la combinación y la integración de diversas ontologías de un mismo dominio, lo que ocurre cuando se unen dos organizaciones diferenciadas, o cuando se pretende obtener una ontología de calidad, a partir de las existentes.

3) Herramientas de evaluación de ontologías: aparecen como instrumentos de apoyo que deben asegurar que tanto las ontologías como las tecnologías relacionadas tengan un nivel mínimo de calidad. Para el futuro, este esfuerzo puede también conducir a las certificaciones estandarizadas.

4) Herramientas basadas de la anotación: estas herramientas se han diseñado para permitir a los usuarios insertar informaciones y datos. La mayoría de estas herramientas han aparecido recientemente, junto con la aparición de la idea de la web semántica.

5) Herramientas de almacenamiento y preguntas: son instrumentos que se han creado para permitir utilizar con facilidad las ontologías. La clave está en el intento de que el web se convierta en una auténtica plataforma para transmitir conocimiento.

6) Herramientas de aprendizaje: se utilizan semiautomáticamente para construir ontologías a partir del lenguaje natural.

2.6.5. Lenguajes para la construcción de ontologías

Como hemos presentado en epígrafes anteriores, en la construcción de la Web semántica existen diferentes lenguajes que tributan a la semantización de Internet. Pero existen otros lenguajes que, unidos a los ya vistos, tributan a la construcción y desarrollo de ontologías, cada uno con sus características

específicas, permitiendo niveles de expresividad para describir los elementos que la componen: conceptos, relaciones, funciones, axiomas e instancias.

Una ontología puede ser representada en varios lenguajes. Los mismos comenzaron a surgir a comienzos de 1990 y se basan principalmente en lógicas de primer orden, en marcos (frames) combinados con lógicas de primer orden y lógicas descriptivas (Tallarico, 2003).

A decir de García Jiménez los lenguajes para la construcción de ontologías, son igualmente empleados como lenguajes de marcado de ontologías (García Jiménez, 2004). Se distinguen por su constante evolución, con una progresión muy clara, con implicaciones en la elaboración y desarrollo de diferentes herramientas de representación del conocimiento.

Actualmente hay un gran número de lenguajes que permiten la construcción de ontologías. Algunos de ellos se presentan a continuación:

- CASL (siglas en inglés para Common Algebraic Specification Language) es un lenguaje expresivo para la especificación formal de los requisitos funcionales y de diseño modular del software y funciona como un estándar dentro del área de las especificaciones de software (Bidoit and Mosses, 2004). Ha sido diseñado por el COFI (International Common Framework Initiative) para la especificación y desarrollo algebraico (Astesiano et al., 2002). CASL debe facilitar la interoperabilidad de las muchas herramientas algebraicas de creación de prototipos y verificación existentes. Se aplica a la especificación de las ontologías con el objetivo de obtener mecanismos estructurados y modulados.
- Loom: inicialmente no fue pensado para implementación de ontologías. Está basado en lógicas descriptivas y reglas de producción y provee una clasificación automática de conceptos. Puede representar conceptos, taxonomías de conceptos, relaciones n-arias, funciones, axiomas, y reglas de producción (Tallarico, 2003).

- Common logic (Sowa, 2004) es una especificación para una familia de lenguajes de ontología que pueden ser perfectamente traducidos de unos a otros.
- OBO (Open Biomedical Ontologies³²) es un lenguaje usado para ontologías biológicas y biomédicas. El proyecto OBO es una iniciativa destinada a la construcción y mantenimiento de una colección evolutiva de ontologías interoperables que de forma inequívoca representen los tipos de entidades en la realidad biológica y biomédica. Esto incluye también un proceso de aseguramiento de calidad para todas las ontologías existentes y creadas por parte de la Fundación (OBO Foundry) que soporta su desarrollo. OBO es una especialización de la ontología formal básica. OBO utiliza un lenguaje propietario denominado OBO-EDIT (Smith et al., 2007). Las ontologías OBO son de dominio público por lo tanto pueden ser usadas libremente por cualquier persona o institución (Caldón et al., 2010).
- SHOE: se creó como extensión del HTML. Usa diferentes tags que permite la inserción de ontologías en documentos HTML. Combina marcos y reglas y puede representar conceptos, taxonomías de conceptos, relaciones n-arias, instancias y reglas de deducción, que son usadas por su motor de inferencias para obtener nuevo conocimiento (Luke and Heflin, 2000).
- XOL: surge a partir de la conversión a XML de un subconjunto pequeño de primitivas del protocolo OKBC (Open Knowledge Base Connectivity) (Tallarico, 2003). Es muy restrictivo y representa conceptos, taxonomía de conceptos y relaciones binarias. No provee un mecanismo de inferencia. Su objetivo es el intercambio de ontologías en dominios biomédicos (Karp et al., 1999).
- OWL: Como ya se había señalado con anterioridad OWL es un estándar. Se encuentra en un nivel más avanzado con respecto a otros lenguajes. Provee de mecanismos de inferencias para determinar la clase del objeto a partir de sus propiedades (Chiaro et al., 2005). Es un lenguaje para hacer

³² Open Biomedical Ontologies <http://www.obofoundry.org/>

declaraciones ontológicas, desarrollado como un seguimiento de RDF y RDFs. Todos sus elementos (clases, propiedades e individuos) son definidos como RDF e identificados por URIs. Es un estándar para la codificación e intercambio de ontologías, basado en lógicas descriptivas que favorecen el desarrollo de aplicaciones semánticas.

OWL posee un conjunto de operadores expresivos para la descripción de conceptos, incluidos los operadores booleanos (intersección, unión y complementar), además de cuantificadores explícitos de propiedades y relaciones; la capacidad de especificar características de las propiedades, tales como la transitividad o dominios y rangos; una bien definida semántica (Bechhofer, 2009).

- OIL (Ontology Interference Layer u Ontology Interchange Language) es el primer lenguaje de representación de ontologías basado en estándares W3C. Tiene sintaxis XML y está definido como una extensión de RDFS. Fue desarrollado por Horrocks y Fensel en el marco del proyecto On-To-Knowledge y el proyecto IBROW(Horrocks et al., 2000). El modelo utilizado por este lenguaje para la representación del conocimiento lo ha obtenido, por una parte de la Lógica Descriptiva (declaración de axiomas o reglas) y por otra, de los sistemas basados en frames (taxonomía de clases y atributos). Permite definir un vocabulario al cual se le asocia un significado que es entendido a nivel de máquina, aunque carece de mecanismos para expresar negociaciones o disyunciones (Chiaro et al., 2005).
- DAML+OIL: La iniciativa DAML+OIL, apunta a proporcionar un lenguaje y un conjunto de herramientas que habiliten la transformación de la Web: de una plataforma que representa información a una plataforma que entienda y razone, que incluso pueda soportar una semántica declarativa, en la que el significado de las expresiones en una representación puede ser entendido sin necesidad de recurrir a un intérprete para su manipulación (Chiaro et al., 2005).

2.6.6. Principales ontologías reutilizables

Las ontologías reutilizables, tras el auge y desarrollo de proyectos para la implementación de la Web semántica, cobran especial importancia en tareas como la integración de la información, la interoperabilidad a nivel de conocimientos y el desarrollo de base de conocimientos. Por lo general hacen uso de la Web para permitir un acceso amplio y proporcionar a los usuarios la capacidad de publicar, buscar, crear y editar ontologías.

Existe una gran variedad de ontologías publicadas, especialmente en el área de las ciencias médicas y biológicas, pero también en el campo de la lingüística, la geografía, negocios y la documentación. Wikipedia (2016) reconoce algunas de ellas:

- BabelNet, una muy grande ontología y red semántica multilingüe, con léxico en muchos lenguajes.
- Basic Formal Ontology, ontología diseñada para soportar investigaciones científicas.
- BioPAX, ontología para el intercambio y la interoperabilidad de los datos de los caminos biológicos (proceso celular)
- BMO, Ontología de modelo basado en análisis de ontologías interpretadas y literatura de modelo de negocios.
- CContology (Customer Complaint Ontology), ontología de negocios que soporta la administración de las quejas de los clientes online.
- COSMO, Foundation Ontology (la actual versión está en OWL) que está diseñada para contener representación de todos los conceptos primitivos necesitados para lógicamente especificar el significado de cualquier dominio. Es un intento de servir como ontología básica que pueda ser usada para el entendimiento entre las representaciones en otras ontologías o bases de datos. Comenzó como una mezcla de elementos básicos de las ontologías OpenCyc y SUMO, y ha sido suplementado con otros elementos de ontologías así como incluir representaciones de todas las palabras en el diccionario Longman definiendo vocabulario.

- Disease Ontology, diseñado para facilitar el reconocimiento de enfermedades y condiciones asociadas a un código médico en particular.
- DOLCE, una ontología descriptiva para ingeniería lingüística y cognitiva
- Foundational Model of Anatomy, ontología para la anatomía humana.
- Gene Ontology para genomas.
- Gellish English dictionary, una ontología que incluye un diccionario y una taxonomía que incluyen ontologías que se centran en las aplicaciones de la industria negocios en ingeniería, tecnología, etc.
- GOLD, General Ontology para Descripción Lingüística.
- GUM (Generalized Upper Model), una ontología lingüísticamente motivada por la intervención entre sistemas clientes y tecnologías de lenguaje natural.
- Linkbase, una representación formal de un dominio biomédico, fundado sobre Basic Formal Ontology.
- NCBO Bioportal, Ontologías biomédicas y biológicas con herramientas para la búsqueda, navegación y visualización.
- Ontologías NIFSTD de Neuroscience Information Framework: un conjunto modular de ontologías de dominio relacionado con la neurociencia.
- OBO-Edit, un buscador de ontología mayoritariamente de Ontologías biológicas y biomédicas.
- OBO Foundry, una suite de referencias ontológicas interoperables en biología y medicina.
- OMNIBUS Ontology, una ontología de aprendizaje, instrucción y diseño instruido.
- Ontology for Biomedical Investigations, un acceso abierto, ontología integrada para la descripción de investigaciones biológicas y clínicas.
- Plant Ontology para estructuras de plantas y estados de crecimiento/desarrollo, etc.

- PRO, the Protein Ontology of the Protein Information Resource, Georgetown University.
- RXNO Ontology, para nombres de reacciones en química.
- SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms).
- Systems Biology Ontology (SBO), para modelos computacionales en Biología.
- Uberon, representa estructuras anatómicas de animal.
- UMBEL, una estructura de referencia ligera de 20,000 temas, conceptos, clases y sus relaciones derivadas de OpenCyc.
- WordNet, un Sistema de referencia léxica.

Existen otras que por sus niveles de alcance y generalización pueden ser aplicadas a cualquier área de conocimiento, y han sido de gran utilidad en la elaboración del presente proyecto. Por dichas razones se detallan a continuación.

2.6.6.1. SKOS (<http://www.w3.org/2004/skos/core>)

SKOS, por sus siglas en inglés Simple Knowledge Organization System (Sistema Simple de Organización del Conocimiento) es un estándar de intercambio de datos (Miles and Bechhofer, 2009). Fue creado por Alistair Millas y Sean Bechhofer y publicado bajo el grupo W3C. (Ver fig 2.22.)

Metadata	
URI	http://www.w3.org/2004/02/skos/core
Namespace	http://www.w3.org/2004/02/skos/core#
isDefinedBy	http://www.w3.org/2009/08/skos-reference/skos.rdf
homepage	http://www.w3.org/2009/08/skos-reference/skos.html
Description	The Simple Knowledge Organization System (SKOS) is a common data model for sharing and linking knowledge organization systems via the Semantic Web. @en
Language	English en
Creator	Alistair Miles http://uri.net/aliman Sean Bechhofer https://plus.google.com/117822622810723317855
Publisher	W3C http://www.w3.org/data#W3C
Comment	(2013-06-06) Bernard Vatant : One of the most used and reused vocabularies both in LOV and LOD, SKOS has had the adoption curve it deserved! (2014-06-06) Bernard Vatant : SKOS is ten years old :) (2015-07-11) Ghislain Atemezing : Annual review done! Added the editors of the W3C document at http://www.w3.org/2009/08/skos-reference/skos.html

Fig. 2.22. Metadatos identificativos de SKOS (fuente: <http://lov.okfn.org/dataset/lov/vocabs/skos>)

El origen del proyecto se remonta a la elaboración de un tesoro de actividades dentro del proyecto SWAD-Europe. La primera versión de SKOS se presentó en el año 2003 (Cavieres Abarca et al., 2010) y posteriormente socializada en su versión SKOS Core 1.0. (Ver Fig.2.23)

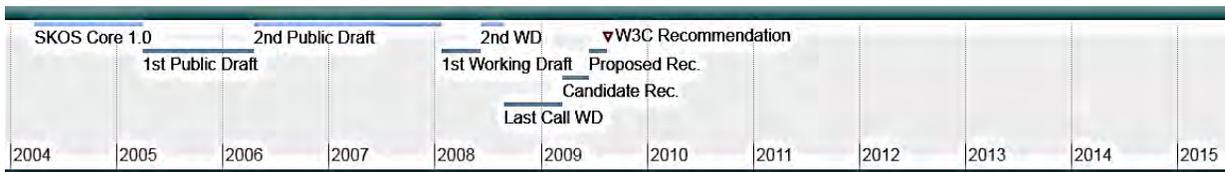


Fig. 2.23. Evolución de las versiones de SKOS (fuente: <http://lov.okfn.org/dataset/lov/vocabs/skos>)

SKOS es una iniciativa del W3C en forma de aplicación de RDF que proporciona un modelo para articular la estructura básica y el contenido de sistemas construidos en pos de la organización del conocimiento y esquemas de

conceptos como tesauros, esquemas de clasificación, listas de encabezamientos de materia, taxonomías, folksonomías, aplicados en sistemas modernos y tradicionales de información, y que sus datos sean legibles por máquinas, así como otros tipos de vocabulario controlado.

Se reconoce el modelo de datos SKOS como una ontología definida con OWL Full (Cavieres Abarca et al., 2010). Posee 4 clases y 28 propiedades. La figura 2.24 muestra la modelación de sus principales clases.

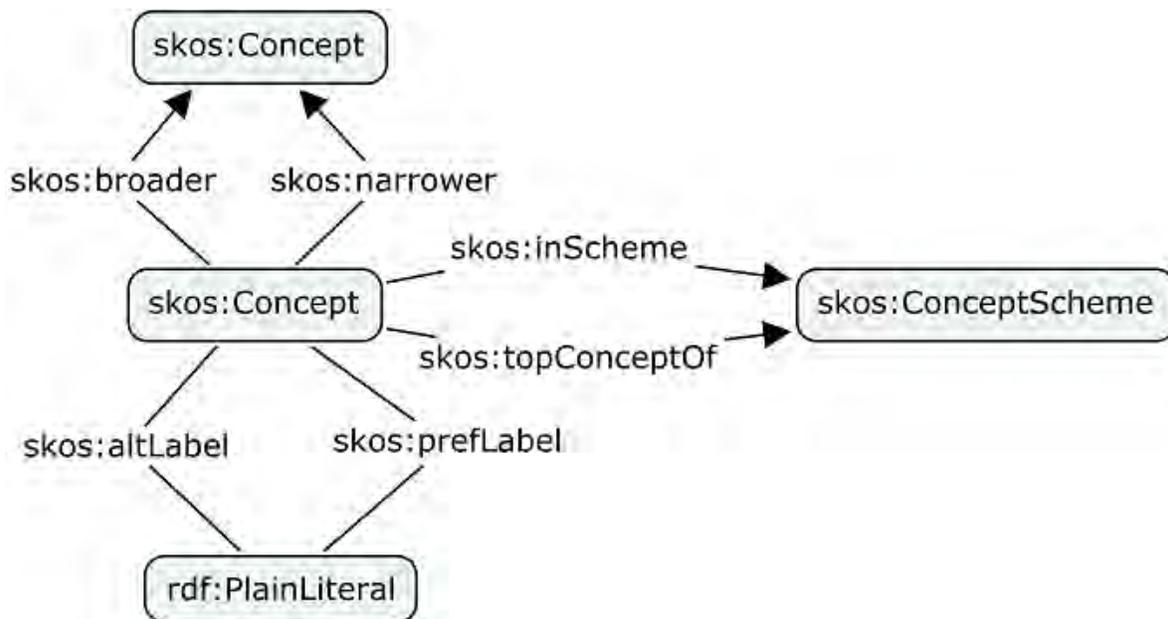


Fig. 2.24. Clases de la ontología SKOS (fuente: <https://tw.rpi.edu/web/project/CMSPV/KeyConcepts>)

El elemento fundamental del vocabulario de SKOS es el *concepto* y lo introduce con la clase `skos:Concept`, que permite asumir un recurso dado como concepto idea, noción o unidad de pensamiento. De este se derivan `skos:ConceptScheme` (se puede ver como una agregación de uno o más conceptos de SKOS), `skos:inScheme` (se utiliza generalmente para describir la relación que un `skos:Concept` "pertenece a" un `skos:ConceptScheme`), `skos:hasTopConcept` (es una subpropiedad de `skos:ConceptScheme`).

SKOS proporciona tres propiedades para adjuntar etiquetas a los recursos conceptuales a través de: *skos:prefLabel*³³(para consignar preferencia sobre una etiqueta), *skos:altLabel*³⁴ (para establecer una etiqueta alternativa) y *skos:hiddenLabel*³⁵. Son todas sub-propiedades de *rdfs:label* y se utilizan para enlazar un *skos:Concept* a un RDF literal, que es una cadena de caracteres (por ejemplo, "love"), combinada con una etiqueta de idioma opcional (por ejemplo, "en-US") (Isaac and Summers, 2009).

Por otra parte dispone de tres propiedades estándar para establecer relaciones entre conceptos: *skos:broader*, *skos:narrower* y *skos:related*.

Las propiedades *skos:broader*³⁶, *skos:narrower*³⁷ se utilizan para permitir la representación de vínculos jerárquicos, como la relación entre un género y sus especies más específicos, o, dependiendo de las interpretaciones, la relación entre un todo y sus partes. Es inverso el uno del otro. De acuerdo con el

³³La propiedad *skos:prefLabel* se utiliza para establecer marcas léxicas. Hace que sea posible asignar una etiqueta léxica preferente a un recurso

³⁴La propiedad *skos:altLabel* se utiliza para establecer etiquetas léxicas alternativas. Permite asignar una etiqueta léxica alternativa a un concepto, cuando los sinónimos necesitan ser representados. Por ejemplo:

```
animales rdf: type skos: Concept;
skos: prefLabel "animales"@es;
skos: altLabel "criaturas"@es;
skos: prefLabel "animaux"@fr;
skos: altLabel "criaturas"@fr.
```

Puede utilizarse además para establecer relaciones de cercanía conceptual, abreviaturas, acrónimos, conceptos agregados a partir de nociones más especializadas y no han sido introducidos explícitamente en el esquema de organización del conocimiento (aunque este último no es recomendado)

³⁵ La propiedad *skos:hiddenLabel* se utiliza para establecer etiquetas léxicas ocultas. Es una etiqueta léxica de un recurso, donde la cadena de caracteres para ser accesible a las aplicaciones que realizan operaciones de indexación y búsqueda basadas en texto no es visible como etiqueta. Habitualmente se utilizan para incluir en la herramienta de organización de conocimiento variantes mal escritas de otras etiquetas léxicas.

³⁶ Se utiliza que un concepto es semánticamente más amplio o general que otro.

³⁷ Se utiliza para marcar que un concepto es semánticamente más específico que otro.

modelo de datos de SKOS, siempre que un concepto es más amplio que otro concepto, entonces hay un concepto más específico que el otro.

La propiedad ***skos:related*** permite la representación de vínculos asociativos (no jerárquicos), como la relación entre un tipo de evento y una categoría de entidades que normalmente participan en él y para representar enlaces parte-todo que no están destinados como las relaciones jerárquicas.

SKOS ofrece la propiedad ***skos:note*** para fines generales de documentación. Esta propiedad está más especializada en para adaptarse a tipos más específicos de documentación:

- ***skos:scopeNote***: Ofrece información parcial sobre el significado pretendido de un concepto, especialmente como una indicación de cómo el uso de un concepto es limitada en la práctica de indexación
- ***skos:definition***: Proporciona una explicación completa del significado previsto de un concepto.
- ***skos:example***: Ofrece un ejemplo del uso de un concepto
- ***skos:historyNote***: Describe cambios significativos en el significado o la forma de un concepto
- ***skos:editorialNote*** : Aporta información que es una ayuda a la limpieza de administración, tales como recordatorios de trabajo editorial que queda por hacer, o advertencias en el caso de que podrían hacerse futuros cambios de redacción.
- ***skos:changeNote***: para consignar los cambios en las versiones de las anotaciones hechas sobre los recursos.

Utilizando SKOS los conceptos pueden ser identificados usando URIs, etiquetados con cadenas léxicas en uno o más lenguajes naturales, anotados (con códigos léxicos), documentados con varios tipos de notas, enlazados a otros conceptos y organizados en jerarquías informales y redes de asociación, agregados a esquemas conceptuales, agrupados en colecciones ordenadas y/o etiquetadas, y contrastados con otros conceptos en otros esquemas (Miles and Bechhofer, 2009, Zeng and Zummer, 2009).

Como aplicación del RDF, SKOS expresa sus datos en tripletas, por lo que admite que puedan ser codificados utilizando cualquier sintaxis RDF. Como consecuencia de ello son legibles por máquina. Maneja conceptos para la creación y publicación en la WWW, vinculado con los datos en la Web e integrado en otros esquemas de conceptos (Isaac and Summers, 2009, Miles and Bechhofer, 2009).

El modelo de datos de SKOS proporciona una ruta de migración de bajo costo estándar para portar los sistemas de organización del conocimiento existentes a la Web Semántica. Su uso para publicar esquemas de conceptos hace que sea fácil hacer referencia a los conceptos en las descripciones de recursos en la Web Semántica. Proporciona un lenguaje ligero e intuitivo para desarrollar y compartir nuevos sistemas de organización del conocimiento. Se puede utilizar por sí solo, o en combinación con lenguajes formales de representación del conocimiento, tales como el lenguaje de ontologías Web (OWL).

En la actualidad, se dispone de dos versiones de aplicaciones de SKOS:

- SKOS básico: los recursos conceptuales (conceptos) se pueden identificar usando URIs³⁸, etiquetarlos con cadenas en una o más lenguas naturales, documentado con varios tipos de notas, semánticamente relacionados entre sí mediante estructuras jerárquicas informales y redes asociativas, y agregarlos a esquemas de conceptos distintos.
- SKOS avanzado: los recursos conceptuales se pueden asignar a los recursos conceptuales en otros esquemas y se agrupan en colecciones etiquetadas u ordenadas. Las relaciones pueden ser especificadas entre las etiquetas de concepto.

SKOS como vocabulario puede ser desarrollado para adaptarse a las necesidades de las comunidades de usuarios concretas o combinadas con otros vocabularios de modelado (ver fig.2.25).

³⁸ Permitiendo que cualquiera pueda referirse a ellos de forma inequívoca de cualquier contexto.

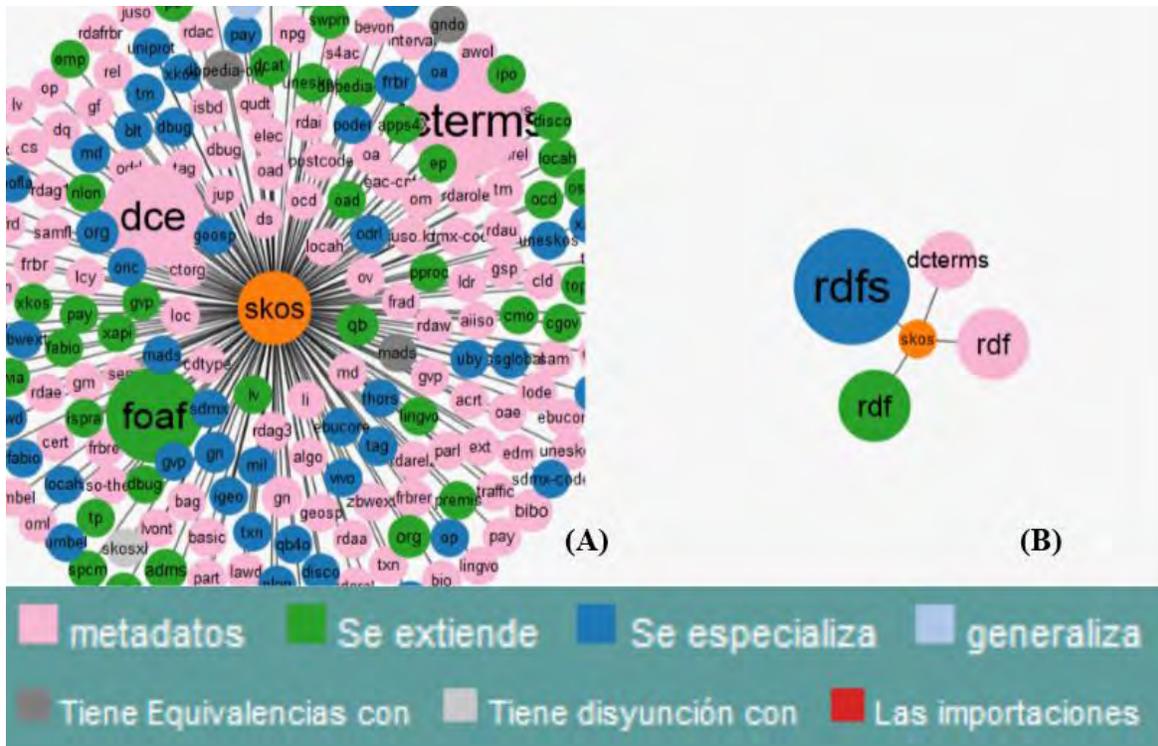


Fig. 2.25. Relación de SKOS con otros vocabularios: (A) 204 vocabularios referenciados por SKOS como lenguajes entrantes; (B) 4 vocabularios referenciados en SKOS como lenguajes salientes. Fuente: <http://lov.okfn.org/dataset/lov/vocabs/skos>

Según declaración de W3C (2012)³⁹ los principales documentos que rigen el funcionamiento y desarrollo de SKOS son:

(Tutoriales)

- *SKOS (Simple Knowledge Organization System) - Dublin Core 2011 tutorial, Antoine Isaac, 21.09.2011*. <http://dcevents.dublincore.org/index.php/IntConf/dc-2011/paper/view/69>

Presenta las diversas características de SKOS, que ilustra la forma en que se pueden utilizar para representar datos. Discute herramientas y métodos para la creación y publicación de SKOS (W3C, 2012b).

³⁹Última actualización: 2012/01/01 23:19:23 por Autor: **Antoine Isaac**. Disponible en: <https://www.w3.org/2004/02/skos/references>

- ***The Simple Knowledge Organization System (SKOS) in the Context of Semantic Web Deployment, Alistair Miles, 2008-07-21.***

http://www.iskouk.org/presentations/miles_web_and_skos_200807.pdf

Presentación "Vocabularios compartir en la web a través de SKOS" en el Evento ISKO del Reino Unido. Presenta las principales motivaciones y razones detrás de la norma SKOS, con ejemplos de aplicación (W3C, 2012b).

- ***On practical aspects of enhancing semantic interoperability using SKOS and KOS alignment, Antoine Isaac, 2008-07-21.***

http://www.iskouk.org/presentations/isaac_21072008.pdf

Presenta el modelo SKOS detallado (prácticas). Se considera un complemento útil para los documentos anteriormente presentados al Taller ISKO Reino Unido. Aborda cómo el uso de SKOS mejora la interoperabilidad de vocabularios conceptuales (W3C, 2012b).

(Documentos introductorios)

- ***Simple Knowledge Organization and the Semantic Web, Alistair Miles, 2005-10.*** <http://isegserv.itd.rl.ac.uk/public/skos/press/Bliss2005/skos-bliss-bulletin-2006.pdf>

Presentación realizada a la Conferencia Anual de la Clasificación de Bliss 2005. Breve artículo para el boletín anual de la Asociación de Clasificación Bliss, para acompañar a la conferencia (W3C, 2012b).

- ***SKOS Core: Simple Knowledge Organization for the Web, Alistair Miles, Brian Matthews y Michael Wilson, 2005-09-12.***

<http://isegserv.itd.rl.ac.uk/cvs-public/~checkout~/skos/press/dc2005/presentation.ppt>

Introducción a SKOS Core para la Conferencia Internacional sobre Dublin Core y Aplicaciones de Metadatos de 2005, que cubre SKOS Core principales características, el estado y el desarrollo (W3C, 2012b).

(Presentaciones generales)

- **Introduction to SKOS - webcast, Bob DuCharme, conferencia SemTech, 2011-10.** http://semanticweb.com/webcast-introduction-to-skos-with-bob-ducharme_b23802

Una visión general de SKOS para la conferencia de Tecnología Semántica orientado a los negocios (W3C, 2012b).

- **SKOS: Past, Present and Future, Sean Bechhofer, 2010-06-03.** <http://www.eswc2010.org/program-menu/keynote-speakers/155-sean-bechhofer?format=pdf>

Conversatorio de apertura en la Conferencia Extended Semantic Web, que ofrece una visión completa de las características de SKOS y su proceso de diseño, aplicaciones de SKOS (W3C, 2012b).

- **Towards the Management of Meaning, Alistair Miles, 2006-12.** <http://isegserv.itd.rl.ac.uk/public/skos/press/oxswig20061206.pdf>

Una presentación entregada a la Web Semántica Grupo de Interés Oxfordshire, dando una breve reseña de la labor SWDWG. Expone desafíos económicos y prácticos en la aplicación de los vocabularios estructurados controlados (W3C, 2012b).

- **Taxonomies and the Semantic Web, Alistair Miles, 2006-02.** <http://isegserv.itd.rl.ac.uk/public/skos/press/cistrana200602/taxonomies-semanticweb.ppt>

Presentación realizada en el taller sobre "CISTRANA Portales para la difusión de información y taxonomías para la clasificación". Aborda la aplicación de las tecnologías de la Web Semántica a las taxonomías de negocios y hasta portales web (W3C, 2012b).

En dicha página puede encontrarse referencia a la publicación de experiencias de corte técnico, referidas al uso y aplicación de SKOS.

2.6.6.2. BIBO (<http://bibliontology.com/>)

BIBO, siglas de Bibliographic Ontology (en español, Ontología bibliográfica), es una ontología para la web semántica. Se utiliza para describir entidades bibliográficas como libros o revistas en la web semántica en RDF. Creado por Bruce D’Arcus y Frédéric Giasson (ver fig. 2.26) y es publicada en 2008 por W3C (Ver fig.2.27).

Metadata	
URI	http://purl.org/ontology/bibo/
Namespace	http://purl.org/ontology/bibo/
Description	The Bibliographic Ontology Specification provides main concepts and properties for describing citations and bibliographic references (i.e. quotes, books, articles, etc) on the Semantic Web. @en
Language	English en
Creator	<div style="display: flex; justify-content: space-between;"> <div style="border: 1px solid black; padding: 2px;">Bruce D’Arcus http://google.com/+BruceDArcus</div> <div style="border: 1px solid black; padding: 2px;">Frédéric Giasson https://plus.google.com/113451969262574338906</div> </div>
Comment	(2013-06-02) Ghislain Atezing: THE ontology to look at first when dealing with bibliographical records on the semantic web. (2014-06-03) Bernard Vatant: Annual review OK (2015-06-02) Bernard Vatant: Annual review - no change

Fig. 2.26. Metadatos identificativos de BIBO (fuente: <http://lov.okfn.org/dataset/lov/vocabs/bibo>)

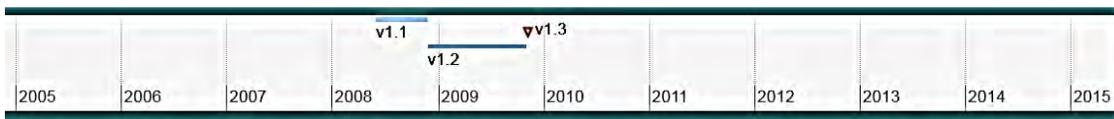


Fig. 2.27. Evolución de las versiones de BIBO (fuente: <http://lov.okfn.org/dataset/lov/vocabs/bibo>)

BIBO puede ser utilizada como una ontología citación, como una clasificación ontología documento, o como una forma de describir cualquier tipo de documento en RDF. Se ha inspirado en diversos formatos de metadatos de descripción de documentos, lo que permite su implementación como base

común para la conversión de otras fuentes de datos bibliográficas. (Ver fig. 2.28).

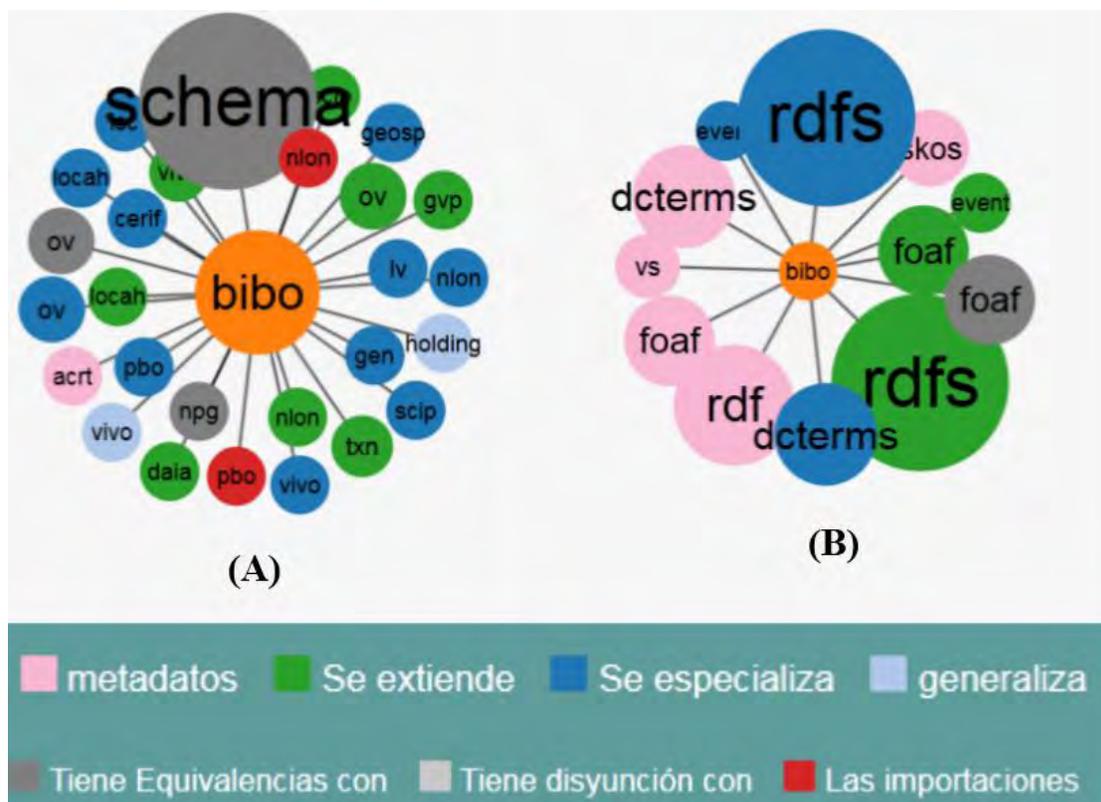


Fig. 2.28. Relación de BIBO con otros vocabularios: (A) 28 vocabularios referenciados por BIBO como lenguajes entrantes y (B) 12 vocabularios referenciados en BIBO como lenguajes salientes (fuente: <http://lov.okfn.org/dataset/lov/vocabs/bibo>)

BIBO proporciona principales conceptos y propiedades para describir las citas y referencias bibliográficas (es decir, citas, libros, artículos, etc.) sobre la Web Semántica. Dispone de aproximadamente 58 clases (ver fig. 2.29) y 67 propiedades.

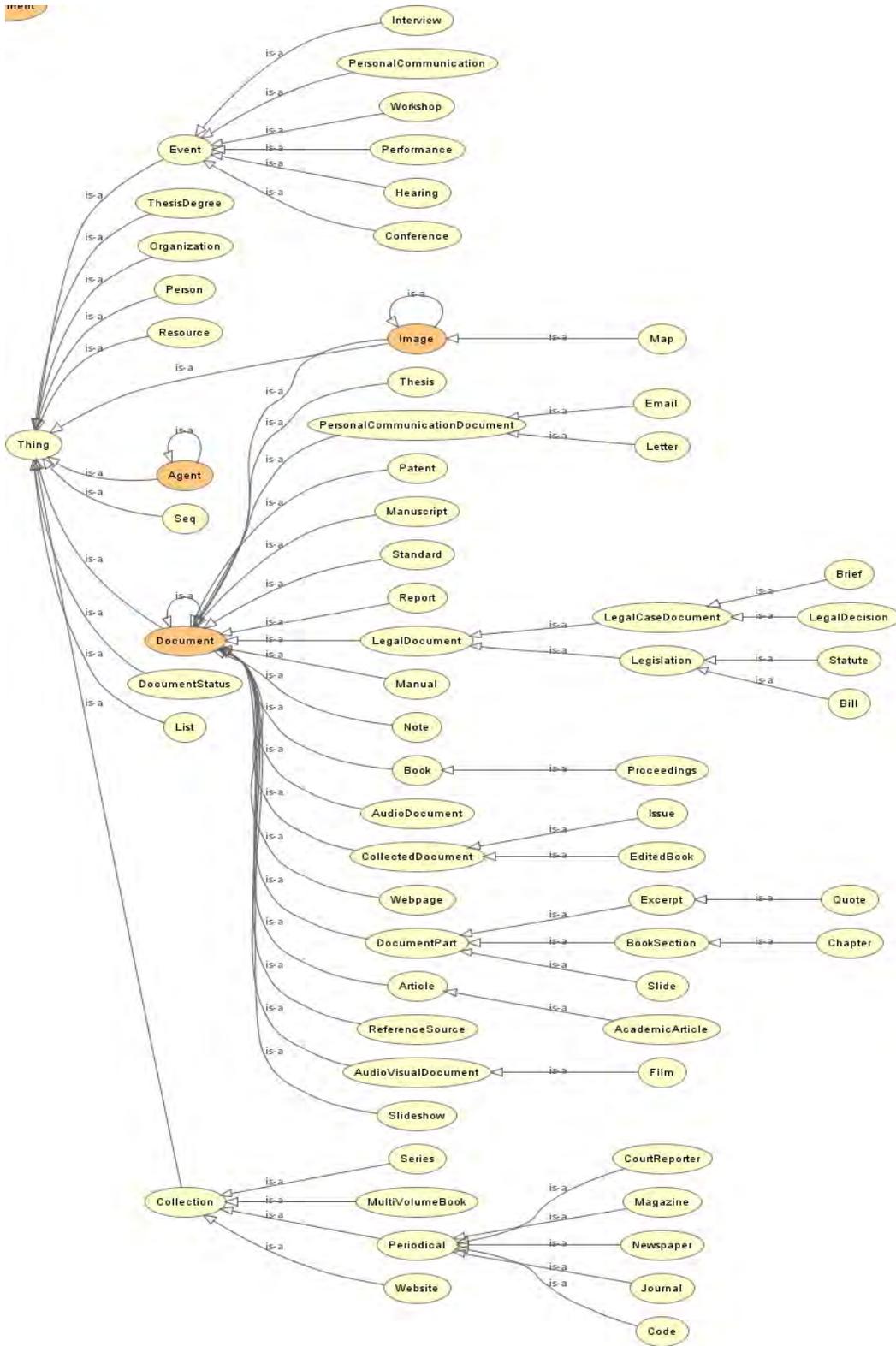


Fig. 2.29. Clases de la ontología BIBO. (Entidades y propiedades tales como identificadores no se muestran en este mapa) fuente: <https://wiki.surfnet.nl/display/standards/info-eu-repo>

2.6.6.3. VIVO (<http://vivoweb.org/ontology/core>)

VIVO es una plataforma construida bajo la filosofía de la web semántica. Es de acceso abierto y permite registrar y gestionar datos referidos a la investigación y la docencia en las múltiples disciplinas por medio de perfiles profesionales vinculados e información relacionada como credenciales académicas, redes de trabajo, información sobre publicaciones, proyectos, servicios , entre otras. Facilita la colaboración entre personas en el ámbito interno de las organizaciones y además entre los diferentes sectores.

VIVO fue desarrollado originalmente por la Universidad de Cornell, que luego (a partir de 2009) fue desarrollado por un consorcio financiado por el NIH (Ding et al., 2011) (cinco universidades en Estados Unidos), lo ampliaron como una herramienta capaz de integrar perfiles entre varias instituciones. En 2013, el IICA con ayuda del Departamento de Agricultura de los Estados Unidos y la Universidad de Cornell iniciaron la adaptación de la herramienta al idioma español, inicialmente, y con el tiempo se han hecho adaptaciones para otros idiomas (Ver Fig. 2.30).

Metadata	
URI	http://vivoweb.org/ontology/core
Namespace	http://vivoweb.org/ontology/core#
homepage	http://www.vivoweb.org/download
Description	An ontology of academic and research domain, developed in the framework of the VIVO project @en
Language	<div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; padding: 2px;">Arabic ar</div> <div style="border: 1px solid black; padding: 2px;">English en</div> <div style="border: 1px solid black; padding: 2px;">Spanish es</div> <div style="border: 1px solid black; padding: 2px;">French fr</div> <div style="border: 1px solid black; padding: 2px;">Chinese zh</div> </div>
Comment	<p>(2014-08-12) Bernard Vatant: Uses an obsolete namespace for SKOS</p> <p>(2013-03-21) Bernard Vatant: Imports a local copy of the FAO geopolitical ontology, and a list of specific instances of bibo:DocumentStatus. Also imports a very small vocabulary http://vivoweb.org/ontology/core/vivo-pws-public-1.5.owl, not considered worth including in LOV as a separate entry.</p> <p>(2014-03-24) Bernard Vatant: Record updated with version 1.6. This ontology reuses also many elements in the http://purl.obolibrary.org/obo/ namespace. Vocabularies in this namespace are not included in LOV.</p> <p>(2015-07-31) Ghislain Atemezing: Annual review - OK</p>

Fig. 2.30. Metadatos identificativos de VIVO (fuente: <http://lov.okfn.org/dataset/lov/vocabs/vivo>)

A partir del 2010 se comienzan a publicar sus diferentes versiones (Ver fig. 2.31.) y junto a otras aplicaciones compatibles producen una extensa red de

conocimiento entre instituciones, organizaciones y agencias, las que en sus búsquedas contribuyen al trabajo colaborativo, las sinergias y a la apertura del conocimiento. El software abierto de VIVO (en sus versiones en español e inglés) y sus ontologías están disponibles públicamente en <http://vivoweb.org>, así como los materiales de soporte para implementar, adoptar o desarrollar.



Fig. 2.31. Evolución de las versiones de VIVO (fuente: <http://lov.okfn.org/dataset/lov/vocabs/vivo>)

Actualmente se ha establecido como un proyecto de código abierto con participación de las comunidades de todo el mundo. A finales de 2012, más de 20 países y 50 organizaciones proporcionarán información en formato VIVO en más de un millón de investigadores y el personal de investigación, incluyendo publicaciones, recursos de investigación, eventos, financiación, cursos que se imparten, y otra actividad académica.

VIVO ofrece una ontología expresiva (gestiona RDF, RDFa y OWL) y permite realizar búsquedas federadas para mejorar la recuperación de información. Los perfiles profesionales y sus descripciones pueden ser importados de manera programada de fuentes oficiales tales como registros institucionales, repositorios locales y otras bases de datos bibliográficas.

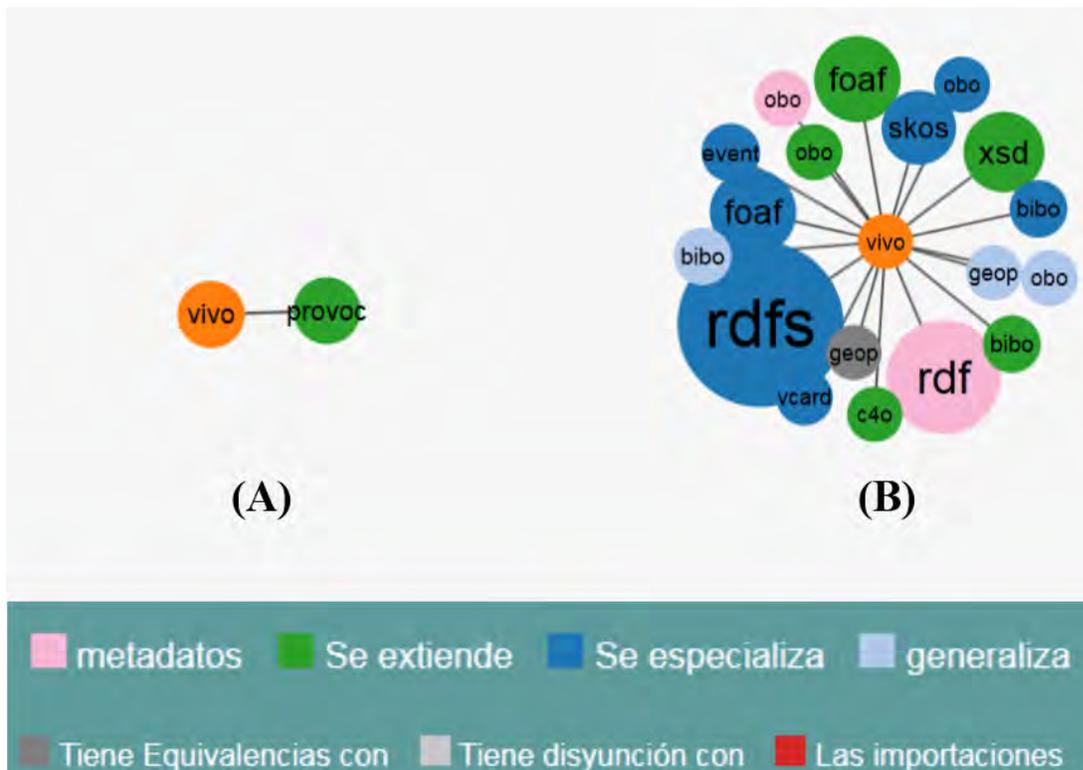


Fig. 2.32. Relación de VIVO con otros lenguajes: (A) 1 vocabularios referenciados por VIVO como lenguajes entrantes y (B) 18 vocabularios referenciados en VIVO como lenguajes salientes (fuente: <http://lov.okfn.org/dataset/lov/vocabs/vivo>)

La ontología VIVO se ha diseñado tomando como foco en el modelado el ámbito científico investigativo, es decir, científicos, publicaciones, recursos, subvenciones, lugares y servicios. Cuenta con 146 clases y 107 propiedades. Incorpora clases de ontologías populares, tales como BIBO⁴⁰, Dublin Core⁴¹, Evento⁴², FOAF⁴³, Geopolitical ontology⁴⁴ y SKOS (Ver fig.2.32 y 2.33).

⁴⁰<http://bibliontology.com/>

⁴¹<http://dublincore.org/documents/dces/>

⁴²<http://motools.sf.net/event/event.html>

⁴³<http://www.foaf-project.org/>

⁴⁴<http://aims.fao.org/aos/geopolitical.owl>

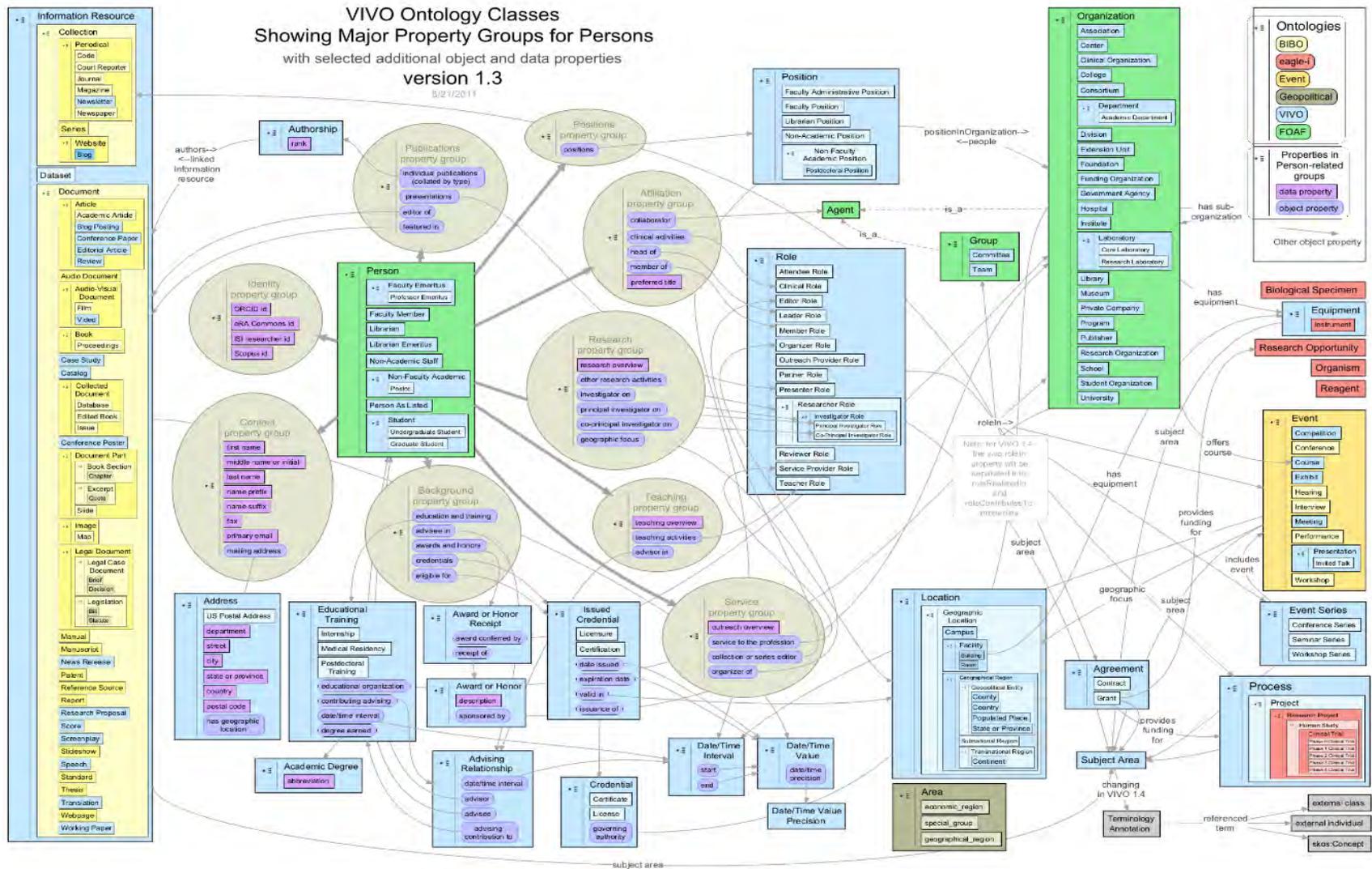


Fig. 2.33. Ontologías integradas en VIVO Ontology y sus clases principales (Ding et al., 2013)

Tabla 2.2. Reseña a las ontologías externas sobre las que hace referencia VIVO (Ding et al., 2011)

External ontologies	namespace	Descriptions	Use in VIVO
Bibliographic ontology	bibo	The Bibliographic Ontology Specification provides main concepts and properties for describing citations and bibliographic references (i.e. quotes, books, articles, etc) on the Semantic Web. ³	Used to model scientific products, mainly publications.
Dublin Core elements & Dublin Core terms	dc	The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description. "core" because its elements are broad and generic, usable for describing a wide range of resources. ⁴	Used to model generic documents
Event Ontology	event	This ontology is centered around the notion of event, seen here as the way by which cognitive agents classify arbitrary time/space regions. ⁵	Used to be the upper class of other specific events
FOAF	foaf	FOAF defines an open, decentralized technology for connecting social Web sites, and the people. ⁶	Used to model scientists, students, staff, etc.
Geopolitical ontology	geopolitical	The geopolitical ontology is a mechanism to describe, manage and exchange data related to geopolitical entities such as countries, territories, regions and other similar areas. ⁷	Used to describe the geopolitical aspects of any entities
Simple Knowledge Organization System (SKOS)	skos	SKOS is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading lists and	Used to represent knowledge organization systems

2.6.6.4. FOAF (<http://www.foaf-project.org/>)

FOAF (Friend of a Friend), es una ontología la gestión de información sobre personas, sus actividades y sus relaciones con otras personas, objetos y la información que utilizan la Web. Creada por Libby Miller y Dan Brickley (ver fig.2.34). Se define como un diccionario de términos, cada uno de los cuales corresponde a una clase o una propiedad (Brickley and Miller, 2014).

Metadata	
URI	http://xmlns.com/foaf/0.1/
Namespace	http://xmlns.com/foaf/0.1/
homepage	http://www.foaf-project.org/
Description	FOAF is a project devoted to linking people and information using the Web. Regardless of whether information is in people's heads, in physical or digital documents, or in the form of factual data, it can be linked. @en
Language	
Creator	<div style="display: flex; justify-content: space-between;"> <div style="background-color: #333; color: white; padding: 2px;"> Libby Miller http://data.semanticweb.org/person/libby-miller </div> <div style="background-color: #333; color: white; padding: 2px;"> Dan Brickley http://google.com/+DanBrickley </div> </div>
Publisher	<div style="background-color: #333; color: white; padding: 2px;"> Dan Brickley http://google.com/+DanBrickley </div>
Comment	<p>(2013-06-04) Bernard Vatant: From the specification : "FOAF has been evolving gradually since its creation in mid-2000. There is now a stable core of classes and properties that will not be changed, beyond modest adjustments to their documentation to track implementation feedback and emerging best practices."</p> <p>(2014-12-16) Bernard Vatant: Annual review OK</p> <p>(2014-01-15) Bernard Vatant: Looking forward for v1.0 :)</p> <p>(2015-12-16) Ghislain Atemezing: Annual review OK</p>

Fig. 2.34. Metadatos identificativos de FOAF (fuente: <http://lov.okfn.org/dataset/lov/vocabs/foaf>)

Integra tres tipos de redes: las redes sociales de colaboración humana, la amistad y de asociación; redes de representación que describen una vista simplificada en términos fácticos, y redes de información que utilizan basada en la Web que une a compartir las descripciones publicadas de forma independiente de este mundo interconectado. FOAF no compite con los sitios Web de orientación social; sino que proporciona un enfoque en el que diferentes sitios pueden decir diferentes partes de la historia más grande, y por el cual los usuarios pueden mantener cierto control sobre su información en un formato no propietario (Brickley and Miller, 2014).

FOAF ha ido evolucionando poco a poco desde su creación a mediados de 2000 (las versiones públicas comienzan a socializarse en 2005, ver fig.2.35). En la actualidad existe un núcleo estable de clases y propiedades que no se van a cambiar, más allá de pequeños ajustes a su documentación para realizar un seguimiento de retroalimentación implementación y mejores prácticas emergentes.



Fig. 2.35. Evolución de las versiones de FOAF (fuente: <http://lov.okfn.org/dataset/lov/vocabs/foaf>)

FOAF contiene varias clases que se muestran en la figura 2.36.



Fig.2.36. Vocabulario FOAF. Esquema basado en la página FOAF Vocabulario Especificación 0.99
 FOAF depende en gran medida de los estándares del W3C trabajo (XML, RDF, OWL). Todos los documentos deben ser FOAF documentos RDF bien formados. El vocabulario FOAF, se logró bajo el estilo de un código abierto o software libre proyecto que como un esfuerzo de la industria de la estandarización. Es bastante simple, pragmático y está diseñado para permitir el despliegue simultáneo. (Ver fig. 2.37)

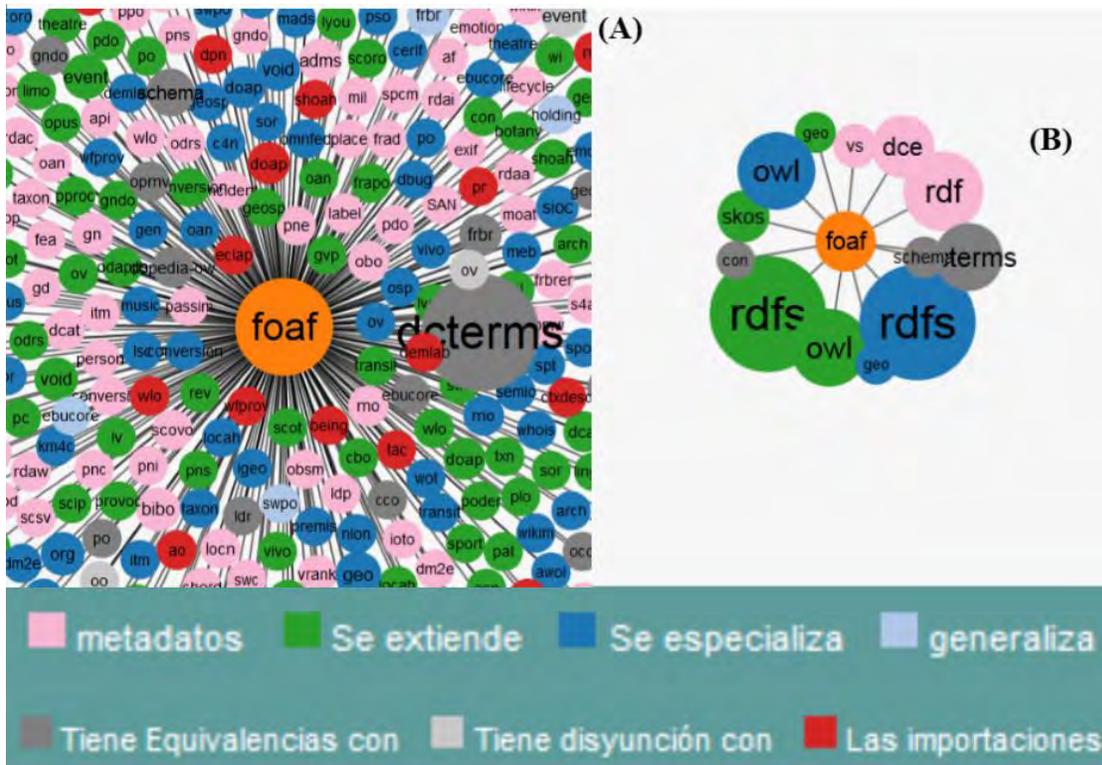


Fig. 2.37. Relación de FOAF con otros vocabularios: (A) 330 vocabularios referenciados por FOAF como lenguajes entrantes y (B) 13 vocabularios referenciados en FOAF como lenguajes salientes (fuente: <http://lov.okfn.org/dataset/lov/vocabs/foaf>)

2.6.6.5. Geopolitical Ontology (<http://aims.fao.org/aos/geopolitical.owl>)

Geopolitical ontology es una ontología que describe información geopolítica creada por Food and Agriculture Organization (FAO) (ver fig. 2.38). La ontología geopolítica incluye nombres en diferentes idiomas; con los códigos de los sistemas (UN, ISO, FAOSTAT, AGROVOC, etc.); facilitando una relación entre territorios (fronteras, grupos de membresía, etc.); y localizando cambios históricos. Además, FAO provee servicios web⁴⁵ de ontología geopolítica y constructor de módulos⁴⁶ para descargarlos en diferentes formatos (RDF, XML, y EXCEL).

⁴⁵<http://www.fao.org/countryprofiles/webservices.asp?lang=en>

⁴⁶<http://www.fao.org/countryprofiles/geoinfo/modulemaker/index.html>

Metadata	
URI	http://aims.fao.org/aos/geopolitical.owl
Namespace	http://aims.fao.org/aos/geopolitical.owl#
Description	FAO's geopolitical ontology version 1.1 was populated with FAO, UN and internationally recognized data sources. @en
Language	<div style="display: flex; justify-content: space-around;"> <div>Arabic ar</div> <div>English en</div> <div>Spanish es</div> <div>French fr</div> <div>Portuguese pt</div> <div>Chinese zh</div> </div>
Publisher	<div style="background-color: #333; color: white; padding: 5px;"> Food and Agriculture Organization http://dbpedia.org/resource/Food_and_Agriculture_Organization </div>
Comment	(2014-06-17) Bernard Vatant : Annual review - no change (2013-06-18) María Poveda-Villalón : Reasoning problems with Protege4-1 using Pellet as reasoner. There is at least one Literal (with value 401) of type Integer that should be String. Another solution is to change the definition of the affected DataType's range. (2015-07-31) Ghislain Atemezing : Annual review - OK

Fig. 2.38. Metadatos identificativos de Geopolitical ontology (fuente: <http://lov.okfn.org/dataset/lov/vocabs/geop>)

Esta ontología fue socializada en 2011 y tributa al manejo de datos geográficos de otras ontologías existentes (Ver fig.2.39).



Fig. 2.39. Relación de Geopolitical ontology con otros vocabularios: (A) 3 vocabularios referenciados por Geopolitical ontology como lenguajes entrantes y (B) 5 vocabularios referenciados en Geopolitical ontology como lenguajes salientes (fuente: <http://lov.okfn.org/dataset/lov/vocabs/geop>)

2.6.6.6. Dublin Core (<http://dublincore.org/documents/dces>)

Dublin Core (DC) es un lenguaje para la representación y manejo de datos, a partir de una propuesta de modelo de datos. Por su complejidad y riqueza ha sido catalogado por algunos autores como una ontología simple para la gestión de información sobre documentos y publicaciones. Es una especificación actualizada de todos los términos de metadatos creado y mantenido por la Iniciativa de Metadatos Dublin Core (DCMI) que incluye propiedades, esquemas de codificación de vocabulario, sintaxis de esquemas de codificación y clases (Ver Fig.2.40).

Metadata	
URI	http://purl.org/dc/terms/
Namespace	http://purl.org/dc/terms/
homepage	http://dublincore.org/documents/dcmi-terms
Description	an up-to-date specification of all metadata terms maintained by the Dublin Core Metadata Initiative, including properties, vocabulary encoding schemes, syntax encoding schemes, and classes. @en
Language	English en
Creator	Dublin Core Metadata Initiative http://purl.org/dc/aboutdcmi#DCMI
Publisher	Dublin Core Metadata Initiative http://purl.org/dc/aboutdcmi#DCMI
Comment	(2013-03-07) Bernard Vatant : Prefix restored to dcterms (2014-03-14) Bernard Vatant : This vocabulary is one of the most used in the LOD cloud, and here to stay, even if the purl redirection is sometimes down, like at the time I write this review. (2015-03-24) Bernard Vatant : Annual review OK (2016-05-10) Ghislain Atemezing : Annual review OK

Fig. 2.40. Metadatos identificativos de Dublin Core (fuente: <http://lov.okfn.org/dataset/lov/vocabs/dcterms>)

DC nace como proyecto a finales de los años 90, pero comienza a tomar auge a partir de 2003 a partir de su publicación y socialización (Ver Fig.2.41).



Fig.2.41. Evolución de las versiones de DC (fuente: <http://lov.okfn.org/dataset/lov/vocabs/dcterms>)

DCMI es una organización abierta apoyo a la innovación en el diseño de los metadatos y las mejores prácticas a través de la ecología de los metadatos. Las actividades de la DCMI incluyen trabajos sobre la arquitectura y el modelado, las discusiones y el trabajo colaborativo en Comunidades DCMI y grupos de trabajo DCMI, globales conferencias, reuniones y talleres, y los esfuerzos educativos para promover la aceptación generalizada de los estándares de metadatos y las mejores prácticas. Mantiene una serie de enlaces y relaciones formales e informales con los organismos de normalización y otras organizaciones de metadatos.

La iniciativa DCMI ha realizado una importante labor para que la propuesta DC se integre como parte esencial de la infraestructura de la Web Semántica y para alcanzar un alto grado de estandarización mediante las normas ISO⁴⁷ 15836-2003 y ANSI⁴⁸/NISO⁴⁹ Z3950.85-2001 (Pastor-Sánchez et al., 2012).

El modelo de abstracción de DCMI tiene como objetivo especificar los componentes y constructores usados en Dublin Core definiendo su naturaleza y describiendo como se combinan entre sí para crear estructuras de información. Este modelo cubre tres aspectos claves:

- Cómo se describen los recursos.
- Cómo se estructuran las propiedades que describen los recursos.
- Cómo se organiza el vocabulario para especificar las propiedades que describen los recursos.

⁴⁷ ISO: International Organization for Standardization. Organización cuyo objetivo es la promoción del desarrollo de normas internacionales de fabricación, comercio y comunicación para todas las ramas industriales a excepción de la eléctrica y la electrónica

⁴⁸ ANSI: American National Standards Institute. Organización encargada de supervisar el desarrollo de estándares en Estados Unidos. Es miembro de la ISO y coordina la interrelación entre estándares internacionales y estándares estadounidenses

⁴⁹ NISO: National Information Standards Organization. Es una organización estadounidense de normalización, que opera en el campo de la publicación, así como en aplicaciones bibliotecarias y bibliográficas. Representa los intereses de ANSI en el Comité Técnico 46 (Información y Documentación) de la ISO.

Dublin Core es un sistema de 15 definiciones semánticas descriptivas que pretenden transmitir un significado semántico a las mismas. Estos son: (Ver tabla 2.3)

Tabla 2.3. Elementos descriptivos de DC (Pastor-Sánchez et al., 2012)

Elemento	Significado
Contributor	Colaborador
Coverage	Covertura
Creator	Creador/Autor
Date	Fecha
Description	Descripción
Format	Formato
Identifier	Identificador
Language	Idioma
Publisher	Editor
Relation	Recurso relacionado
Rights	Derechos de autoría, intelectuales, de explotación, etc.
Source	Fuente original de la que se deriva el recurso descrito
Subject	Tema o asunto del recurso
Title	Título
Type	Naturaleza o género del recurso: colección, imagen, recurso interactivo, servicios, software, sonido, texto, etc.

En general, podemos clasificar estos elementos en tres grupos que indican la clase o el ámbito de la información que se guarda en ellos: elementos relacionados principalmente con el contenido del recurso; elementos relacionados principalmente con el recurso cuando es visto como una propiedad intelectual y elementos relacionados principalmente con la instanciación del recurso. Dentro de cada clasificación encontramos los siguientes elementos:

Contenido:

- **Título:** el nombre dado a un recurso, habitualmente por el autor.
(Etiqueta: DC.Title)
- **Claves:** los temas del recurso. Típicamente, Subject expresará las claves o frases que describen el título o el contenido del recurso. Se

fomentará el uso de vocabularios controlados y de sistemas de clasificación formales. (Etiqueta: DC.Subject)

- **Descripción:** una descripción textual del recurso. Puede ser un resumen en el caso de un documento o una descripción del contenido en el caso de un documento visual. (Etiqueta: DC.Description)
- **Fuente:** secuencia de caracteres usados para identificar unívocamente un trabajo a partir del cual proviene el recurso actual. (Etiqueta: DC.Source)
- **Tipo del Recurso:** la categoría del recurso. Por ejemplo, página personal, romance, poema, diccionario, etc. (Etiqueta: DC.Type)
- **Relación:** es un identificador de un segundo recurso y su relación con el recurso actual. Este elemento permite enlazar los recursos relacionados y las descripciones de los recursos. (Etiqueta: DC.Relation)
- **Cobertura:** es la característica de cobertura espacial y/o temporal del contenido intelectual del recurso. (Etiqueta: DC.Coverage)

Propiedad Intelectual:

- **Autor o Creador:** la persona u organización responsable de la creación del contenido intelectual del recurso. (Etiqueta: DC.Creator)
- **Editor:** la entidad responsable de hacer que el recurso se encuentre disponible en la red en su formato actual. (Etiqueta: DC.Publisher)
- **Otros Colaboradores:** una persona u organización que haya tenido una contribución intelectual significativa, pero que esta sea secundaria en comparación con las de las personas u organizaciones especificadas en el elemento Creator. (Etiqueta: DC.Contributor)
- **Derechos:** son una referencia (por ejemplo, una URL) para una nota sobre derechos de autor, para un servicio de gestión de derechos o para un servicio que dará información sobre términos y condiciones de acceso a un recurso. (Etiqueta: DC.Rights)

Instanciación:

- **Fecha:** una fecha en la cual el recurso se puso a disposición del usuario en su forma actual. (Etiqueta: DC.Date)
- **Formato:** es el formato de datos de un recurso, usado para identificar el software y, posiblemente, el hardware que se necesitaría para mostrar el recurso. (Etiqueta: DC.Format)
- **Identificador del Recurso:** secuencia de caracteres utilizados para identificar unívocamente un recurso. (Etiqueta: DC.Identifier)
- **Lengua:** lengua/s del contenido intelectual del recurso. (Etiqueta: DC.Language)

Por tanto DC como ontología se nutre de diferentes vocabularios (ver fig.2.42)

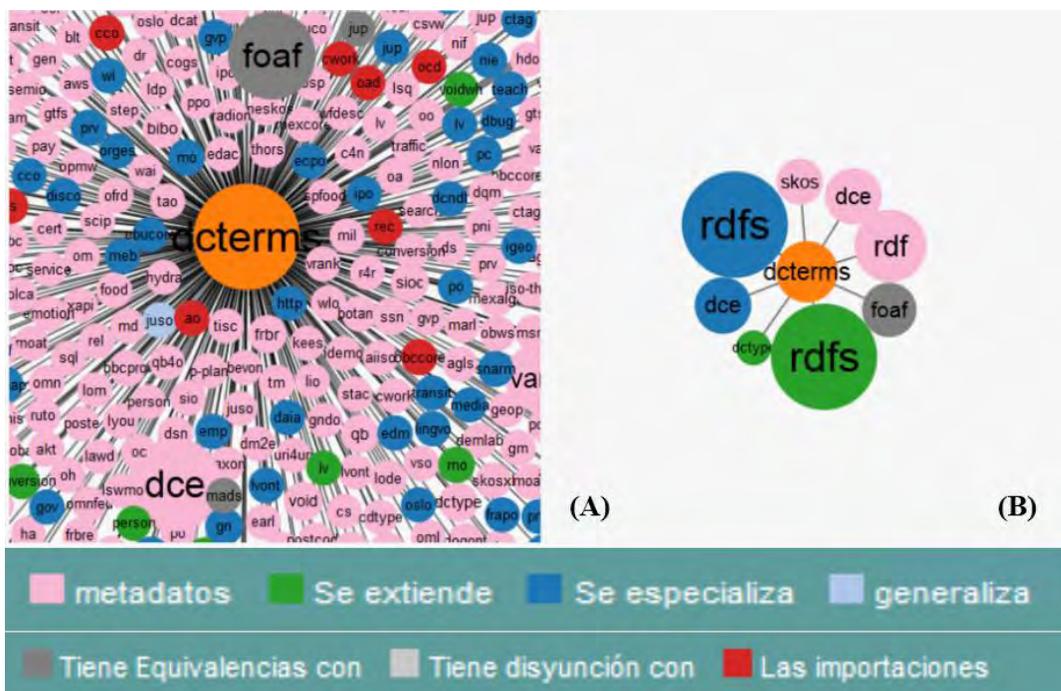


Fig. 2.42. Relación de DC con otros vocabularios: (A) 444 vocabularios referenciados por DC como lenguajes entrantes y (B) 8 vocabularios referenciados en DC como lenguajes salientes (fuente: <http://lov.okfn.org/dataset/lov/vocabs/dcterms>)

2.7. Web semántica y Recuperación de la Información

El fin de la Web semántica es la representación y descripción del contenido semántico de recursos de información para tratar de solventar los problemas

de aplicación de técnicas para hacer más eficientes los procesos de búsqueda y recuperación de la información en la red. Esto ha sido un incentivo para el desarrollo de nuevas técnicas que han ido más allá de la simple extracción de términos.

2.7.1. Recuperación de la información: conceptualización

La expresión "recuperación de información" fue acuñada por Calvin Mooers en 1948. Pese a ser empleada desde el principio de los años 50 su gran difusión a partir de la década siguiente implica una gran variedad de definiciones (Pastor-Sánchez et al., 2012).

Desde entonces varios autores abordaron la conceptualización del término en función, de los avances tecnológicos del contexto social. Entre las más relevantes puede referirse a Salton quien señalaba que "**la recuperación de la información tiene que ver con la representación, almacenamiento, organización y acceso a los ítem de información**" (Salton and McGill, 1983).

Para Croft **la recuperación de información es "el conjunto de tareas mediante las cuales el usuario localiza y accede a los recursos de información que son pertinentes para la resolución del problema planteado. En estas tareas desempeñan un papel fundamental los lenguajes documentales, las técnicas de resumen, la descripción del objeto documental, etc."** (Croft, 1987).

Feather y Storges consideran que la recuperación de información es "**el conjunto de actividades necesarias para hacer disponible la información a una comunidad de usuarios**" (Feather and Sturges, 1997).

Según Pastor Sánchez la recuperación de información no debe limitarse a un proceso en la que el usuario realiza una consulta y el sistema de recuperación de información devuelve un conjunto ordenado de elementos relevantes. Coincide con Salton (1999) en que el problema de la recuperación de información tiene que ver en gran medida con la forma en la que se representa la información. Enfatiza que al trabajar con información documental sin tener en cuenta la estructura semántica de la misma, se estará desarrollando una tarea más cercana a la recuperación datos que a la recuperación de información (Pastor-Sánchez et al., 2012).

El fenómeno de la recuperación de información ha sido abordado como proceso, mecanismo, herramienta, actividad o tarea, pero siempre vinculado a la posibilidad de acceso a la información y las formas más adecuadas de lograrlo garantizando su pertinencia y relevancia ante la necesidad informativa del usuario.

2.7.2. Sistemas de recuperación de la información (SRI) en la Web

Un sistema de recuperación de información (SRI) está diseñado para analizar, procesar y almacenar fuentes de información y recuperar los que coinciden con los requerimientos de un usuario en particular (Chowdhury, 2010). Los esquemas existentes para la consulta y acceso a la información en la Web, garantizan la recuperación de un conjunto limitado de documentos, debido su heterogeneidad (en cuanto a formato, estructura o la calidad de sus contenidos) y al gran número de páginas web existentes. Varios autores (Baeza-Yates and Ribeiro-Neto, 1999, Kuna et al., 2014) distinguen como los principales modelos de SRI que operan sobre internet son: los directorios, los buscadores (o motores de búsqueda) y los meta-buscadores.

Los motores de búsqueda como Google, explotan la estructura de enlaces para la obtención de la relevancia de las páginas recuperadas. La utilidad de correo Gmail ofrece la posibilidad de gestionar etiquetas para clasificar los mensajes de correo del usuario para su posterior búsqueda y recuperación. Se trata de una búsqueda basada en la exploración de una red de conceptos. Esta evolución de los SRI se debe fundamentalmente al carácter dinámico de la Web desde el punto de vista tecnológico, de contenidos y de servicios ofertados (Savoy and Picard, 2001).

La razón fundamental de los problemas asociados a la recuperación de información está mancomunada a la insuficiencia de herramientas que permitan describir de un modo más exacto el contenido semántico de los recursos de información.

Tal y como se distingue en el proceso evolutivo de la web, los actuales sistemas de recuperación de información para este espacio son el resultado de una evolución continua de la propia Web desde su aparición. (Ver Fig.2.43)

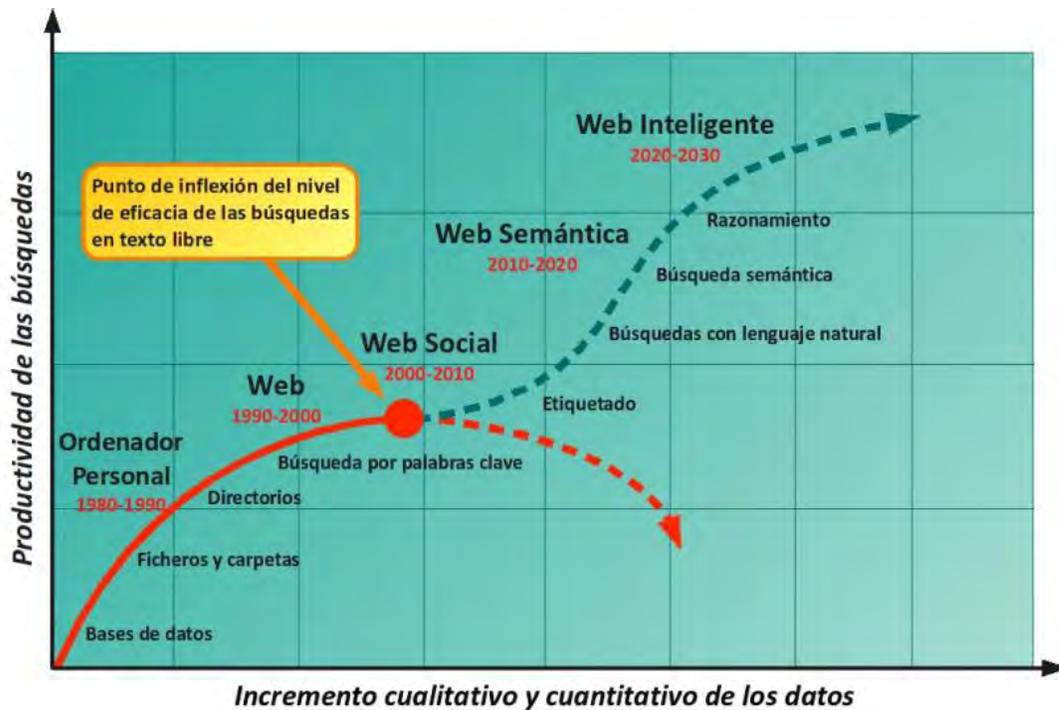


Fig. 0.1Fig.2.43. Eficacia de la búsqueda en texto libre en relación a la evolución de la Web. Traducido y adaptado de (Spivack, 2007) por (Pastor-Sánchez et al., 2012)

En un entorno tan dinámico como la Web el proceso evolutivo de los SRI integra características que actualmente ofrecen de forma individual diversos servicios de búsqueda de información. Los metadatos son el elemento catalizador de la Web Semántica. Según han ido ganando en capacidad descriptiva y ampliando las posibilidades de estructuración e interrelación, se han obtenido nuevas tecnologías que permiten pasar de sistemas de recuperación de información basados en listas de datos hacia aplicaciones que contemplan procesos lógicos para definir y ejecutar tareas cercanas al razonamiento (Davis, 2006, Pastor-Sánchez et al., 2012, Wave, 2006).

También han evolucionado las herramientas y lenguajes usados para codificar metadatos y adaptarse al incremento de las necesidades de estructuración y abstracción de la información. (Ver fig.2.44)

Fig.2.44. Relación entre el incremento de la capacidad de búsqueda y el uso de metadatos. Traducido y adaptado de por (Pastor-Sánchez et al., 2012)

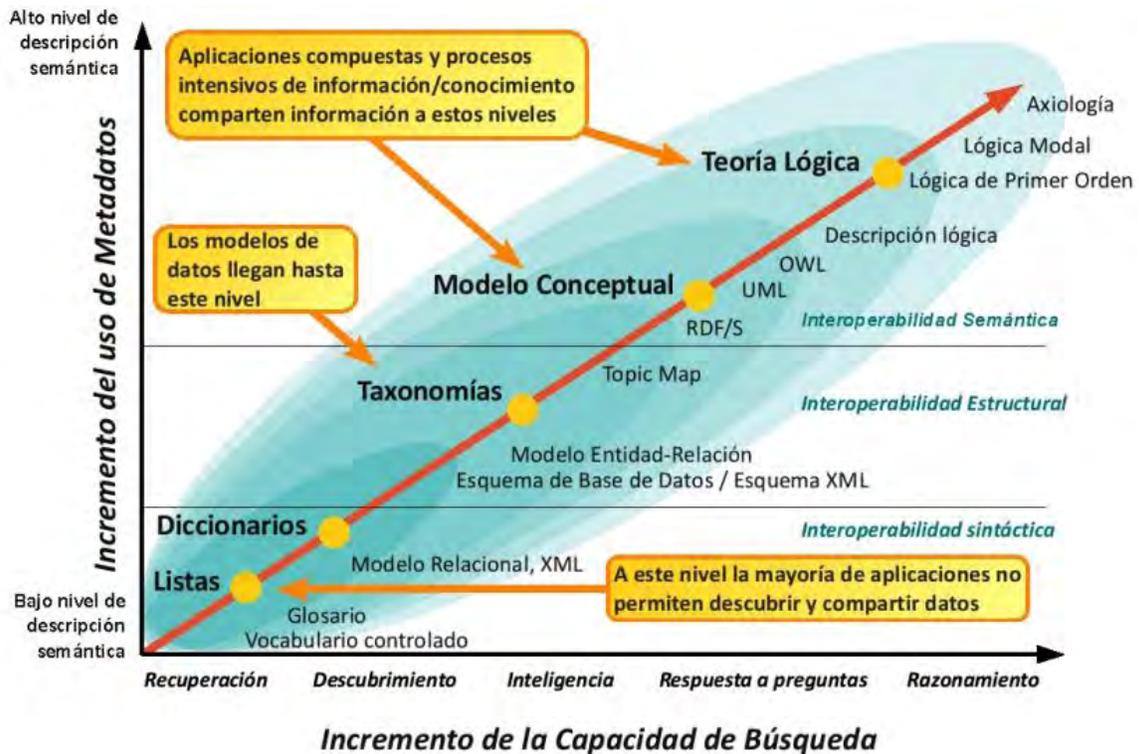


Fig. 2.44. Relación entre el incremento de la capacidad de búsqueda y el uso de metadatos. Traducido y adaptado de (Davis, 2006) por (Pastor-Sánchez et al., 2012)

La Web Semántica necesita el conocimiento en un formato que pueda ser procesado por máquinas. Por tanto, siendo coherente con lo antes expuesto, puede afirmarse que la representación de la información en la Web Semántica se fundamenta en dos conceptos: metadatos y ontologías. Por consiguiente, la búsqueda y recuperación de la información como procesos y los SRI como servicios web, condicionan su calidad a la adecuada gestión de los metadatos y las ontologías.

Los metadatos se han convertido en un núcleo de información estructurada y descrita de forma precisa para poder realizar búsquedas más eficaces. El uso de ontologías complementa las aplicaciones a partir de metadatos para realizar procesos más complejos de forma estructurada, a partir de la descripción de objetos definiendo clases, propiedades, relaciones y axiomas y simplificando la interoperabilidad de diferentes representaciones de datos con herramientas de unificación, traducción y mapeado.

Los SRI basados en ontologías han de permitir, a partir de la representación del conocimiento basado en conceptos (en vez de expresiones o palabras clave) identificar el contenido semántico de los recursos de información independientemente de su representación sintáctica. Han de poseer al menos una doble vertiente (Pastor-Sánchez et al., 2012):

- Semántica: En la que se formaliza, expresa y estructura el contenido con niveles de detalle relativamente flexibles para una descripción de objetos y hechos del mundo real, tanto en sus propiedades como en sus relaciones. (Ver fig. 2.45)
- Pragmática: En donde se definen las posibles aplicaciones de una ontología con métodos de diseño adecuados y teniendo en cuenta el contexto de uso aplicando criterios lógicos de razonamiento, de organización y de rigurosidad, todo ello para obtener dominios de conocimiento, funcionales y aplicables. (Ver fig. 2.45)

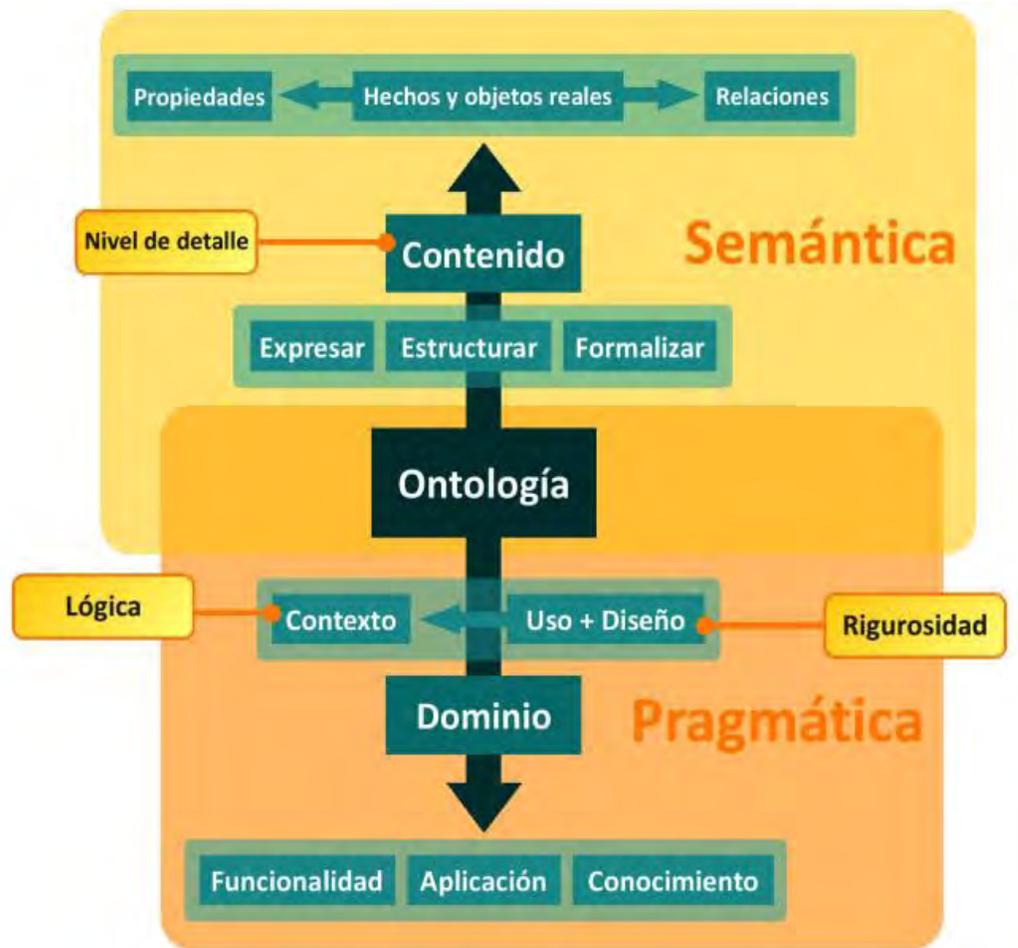


Fig.2.45. Vertiente semántica y Pragmática de una ontología (Pastor-Sánchez et al., 2012)

Cada vez son más las voces que califican la utilización de una ontología para la recuperación de información como un método eficaz, que puede superar a otros métodos en precisión y relevancia (Beltran-Gómez, 2013). Para estos procesos, habitualmente se recurre a los sistemas de expansión⁵⁰ de consultas

⁵⁰ Sistema de expansión: se trata de un proceso por el que se toma la consulta original ingresada por el usuario y se la amplía a partir de la inclusión de diversos términos, obtenidos a partir de fuentes de información externas, que guardan relación con el dominio de la consulta. De esta manera con los términos originales y los obtenidos en forma externa se genera una serie de consultas alternativas, denominadas expansiones. KUNA, H. D., REY, M., PODKOWA, L., MARTINI, E. & SOLONEZEN, L. Year. Expansión de consultas basada en ontologías para un sistema de recuperación de información. *In:* XVI Workshop de Investigadores en Ciencias de la Computación, 2014.

que utilizan ontologías. Este es uno de los componentes que incluye un metabuscador con el objetivo de ampliar el espectro de búsqueda para recuperar documentos de mayor relevancia para el usuario.

Con los sistemas de expansión, desde un punto de vista lingüístico, a partir de una palabra, un sintagma o un conjunto de palabras de la consulta, el sistema buscaría en la ontología otras palabras o sintagmas que expresaran conceptos próximos. Esta proximidad se correspondería, en lingüística y por regla general, con sinónimos o variantes, con hiperónimos o clases de conceptos, y con cohipónimos o conceptos que pertenecen a una misma clase; difícilmente se encuentra en este tipo de sistemas otro tipo de relaciones conceptuales como las de causalidad o de secuencia temporal (Lorente, 2005). A partir de estas consultas, se busca que el SRI adquiriera la capacidad de acceder a nuevos documentos que sean relevantes para el usuario, pudiendo ejecutar en paralelo las consultas a las diferentes fuentes de datos definidas y sus resultados posteriormente ser mezclados y ordenados generando el listado final a ser presentado al usuario (Kuna et al., 2014).

Además de las ontologías, la web semántica propone actividades y tecnologías para minimizar las deficiencias en los procesos de búsqueda y recuperación de la información, es decir, combatir la falta de interoperabilidad, información aislada y dispersa, poca visualización, etc. , a través de la descripción de recursos de la web con representaciones procesables por máquinas, no solo con descriptores explícitos o metadatos, también se emplean los llamados protocolos de sistemas de información documental. Dichos sistemas, a decir de Guajardo (2010), tienen la capacidad de interoperar entre ellos a través de protocolos de transferencia, definen la sintaxis, semántica y sincronización de la comunicación. Entre los protocolos más desarrollados se encuentran (Beltran-Gómez, 2013):

- Z39.50 (Z39.50 Gateway to Library Catalogs) cuenta con funcionalidad muy completa para tratar cuestiones relacionadas con el manejo de sesiones y gestión de conjuntos de resultados (Beltran-Gómez, 2013).

- Simple digital library interoperability protocol (SDLIP) permite a los clientes consultar las fuentes de información en una sintaxis uniforme. El protocolo fue desarrollado en una colaboración entre Stanford, las Universidades de California en Berkeley, y Santa Bárbara, el Centro de Supercomputación de San Diego, y la Biblioteca Digital de California. El protocolo soporta múltiples lenguajes de consulta, y define un formato simple de retorno basado en XML. SDLIP se puede utilizar con CORBA o HTTP (Paepcke et al., 2000).
- OAI (open archives protocol), que aborda las dificultades en la dispersión de recursos y documentos en múltiples depósitos temáticos y/o revistas individuales en la Web implementando un protocolo de recolección de datos sobre cualquier material almacenado en forma electrónica para su posterior publicación en internet (Beltran-Gómez, 2013, Gómez-Dueñas, 2006).
- Starts (Stanford protocol proposal for internet retrieval an search) un protocolo que facilita la tarea de consultar varias fuentes de documentos (Gravano et al., 1997).
- Harvest: a distributed search system. Es un sistema que ofrece un conjunto de herramientas personalizables para recopilar información de diversos repositorios, la construcción de índices de contenido sobre temas específicos, de forma flexible la búsqueda de los índices, ampliamente replicarlos y almacenamiento en caché de objetos a medida que se recuperan a través de Internet. El sistema interactúa con el mosaico y con HTTP, FTP, Gopher y recursos de información. Ofrece altos niveles de personalización (Bowman et al., 1994).

Sin embargo los protocolos Z39.50 y el OAI-PMH se han consolidado internacionalmente. Actualmente, la OAI cuenta con dos vertientes de proveedor de datos (u OAI-PMH), que contiene los metadatos de los recursos a publicar en internet; y proveedor de servicio (u OAI-ORE: Object Reuse and Exchange) que extraen dichos metadatos (utilizando Dublin Core) para brindar el servicio de búsqueda (Beltran-Gómez, 2013).

Finalmente, es válido decir que con la evolución de la Web, las herramientas, formatos y protocolos embebidos en ella, hoy existe una gran variedad de

técnicas disponibles para recuperar la información, tanto por el usuario como para el usuario de la información en sentido general. Este es un proceso de evolución continua al que hay que dedicar especial seguimiento.

2.7.3. SPARQL

El SPARQL es el acrónimo recursivo de SPARQL Protocol and RDF Query Language. Es un lenguaje de consultas sobre grafos RDF estandarizado por la W3C (Tabares and Leiva Mederos, 2016, W3C, 2008, Pastor Sánchez and Díaz Ortuño, 2008). Este estándar de consultas surge no solo con el objetivo de realizar consultas RDF sino que también facilita la recuperación de información en datos provenientes de bases de datos relacionales, estructuras en XML y en JSON. Hay 2 versiones:

- 1.0: Bastante implementada y recomendación W3C.
- 1.1 : Es la recomendación más reciente (marzo 2013)

Las consultas en este lenguaje están basadas en comparación de patrones gráficos, estas son descritas desde la estructura clásica del grafo RDF: sujeto, predicado y objeto, pero con la opción de una variable consulta en lugar de un término RDF (en las posiciones del sujeto, predicado u objeto) (Valencia Castillo, 2008). Las consultas de SPARQL se denominan consultas semánticas y es la base del sistema de búsqueda del CMS semántico y de la web semántica (ver esquemas de la fig 2.8, abordado en el epígrafe referido al tema).

Estas consultas pueden traducirse en los sistemas de búsqueda de las interfaces web para que los usuarios puedan manejar las formas esenciales de este tipo de consulta, cuya eficiencia depende de la construcción de sentencias alojadas en un grafo y de la ejecución de operadores booleanos.

La estructura del lenguaje SPARQL se compone de varios elementos (ver figs. 2.46 y 2.47): (Ducharme, 2013, Bizer and Schultz, 2009)

- **SELECT:** Es una cláusula que permite retornar los valores de un dato específico y construir un parámetro de consulta usando una condición. Identifica las variables que aparecen en los resultados de la consulta

- **PREFIX:** Es la identificación del recurso mediante una URI, lo que constituye la denominación exacta de un tipo de dato o vocabulario donde se van a encriptar los datos.
- **WHERE:** Declara donde seleccionar los datos que debo mostrar en la consulta. Proporciona el patrón de grafo básico para la concordancia con el grafo de datos.

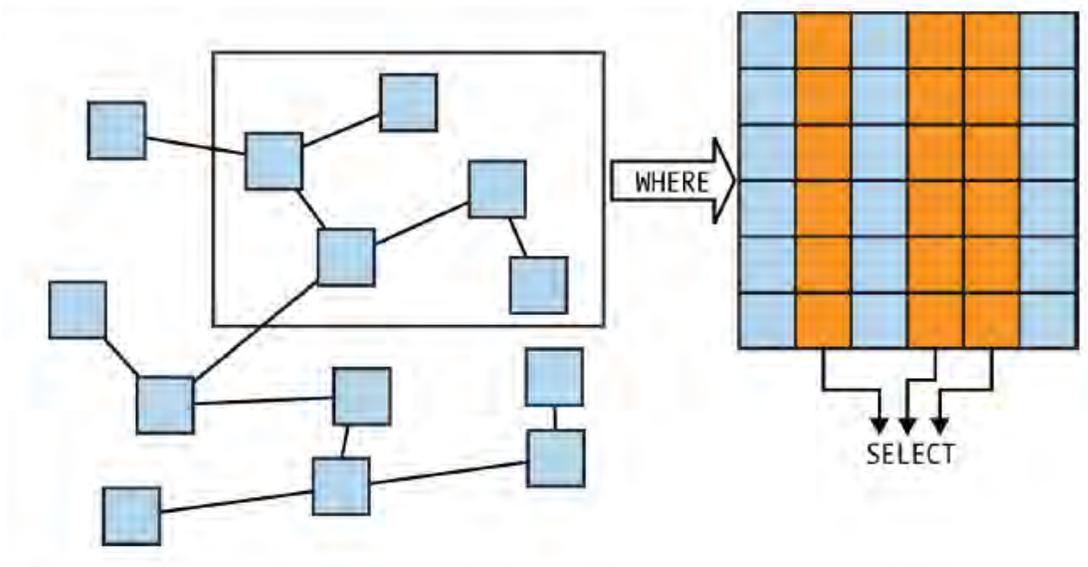


Fig. 2.46. Uso del Where y el Select en las Consultas SPARQL (Ducharme, 2013)

Consulta:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE
{ ?x foaf:name ?name .
  ?x foaf:mbox ?mbox }
```

Resultados de la consulta:

name	mbox
"Johnny Lee Outlaw"	<mailto:jlow@example.com>
"Peter Goodguy"	<mailto:peter@example.org>

Fig. 2.47. Ejemplo de Consultas SPARQL (W3C, 2008)

Las consultas en SPARQL se agrupan en función de los objetivos de la búsqueda y de la precisión que se requiera⁵¹ (Correndo et al., 2010). Según el W3C (2008), SPARQL tiene cuatro formas de consulta. Estas formas de consulta usan las soluciones que concuerdan con el patrón para formar los conjuntos de resultados o grafos RDF. Las formas de consulta son:

- **Select:** Devuelve todo, o un subconjunto de las variables vinculadas en una concordancia con un patrón de búsqueda. (Ver fig.2.46)
- **Construct:** Otra de las consultas más utilizadas en SPARQL. Retornan grafos RDF construido mediante la sustitución de variables en un conjunto de plantillas de tripleta. Esto permite extraer tripletas de una fuente de datos sin modificarlas para luego usarlas o sus valores con el fin de crear nuevos grafos, copiar los originales o hacer una copia modificada de los originales y de esta forma obtener consultas con un rendimiento mayor en el sistema. Ejemplo:

```
construct {?s ?p ?recurso } where {
?s ?p ?recurso .
?s rdfs:label ?label .
FILTER(regex(?label, "José Antonio Senso"))
}
```

Las consultas construct pueden modificarse además para obtener grafos más complejos que de alguna manera faciliten al desarrollador la declaración de grafos de consulta sobre entidades externas.

```
construct {
<http://www.uci.cu/leandro> rdf:type foaf:Person .} where {}
```

- **Ask:** Las consultas de tipo Ask permiten verificar si un patrón de tripletas está presente en un grafo. Esto es muy útil ya que permite verificar la calidad de la información almacenada. Este tipo de consultas devuelve valores booleanos y facilita la búsqueda declarativa en la recuperación de la información. Ejemplo

⁵¹ Es pertinente aclarar que la autora no intentará barrer todo este amplio tema y se centrará solo en las cuestiones que son inherentes a la dimensión de las consultas SPARQL en un CMS

```
ask where {
  <http://es.dbpedia.org/resource/José_Martí> <http://es.dbpedia.org/pro
  perty/web> ?page .
  FILTER (isURI(?page))
}
```

- **Describe:** La forma DESCRIBE devuelve como resultado un único grafo RDF con datos sobre recursos. El patrón de consulta es usado para crear un conjunto de resultados. Toma cada uno de los recursos identificados en una solución, junto con otros recursos directamente nombrados mediante una IRI, y construye un único grafo RDF tomando una "descripción" que puede provenir de cualquier información disponible, incluyendo el conjunto destino de datos RDF objeto.. La sintaxis DESCRIBE es una abreviación que describe todas las variables en una consulta. Obtienen una descripción de un recurso determinado en base a la información almacenada en una fuente de datos determinada.

Ejemplo: URIs explícitas

```
DESCRIBE <http://es.dbpedia.org/resource/José_Martí>
```

Ejemplo: identificación de un recurso

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
DESCRIBE ?x
WHERE { ?x foaf:name "Alice" }
```

Tanto SPARQL es cada vez más utilizado para la creación de estrategias de acceso a la información. Aplicaciones que usan gráficos RDF como FOAF, son accedidos a través de consultas SPARQL en una variedad de plataformas y lenguajes, como Protégé. Este protocolo se presenta como la aplicación de la tecnología semántica a otras áreas como las bases de datos, la ingeniería del software, la gestión del conocimiento, que está cambiando la forma de interactuar entre seres humanos y las máquinas.

2.7.4. Datos Enlazados

- Usar URIs HTTP: Las URIs no son URLs (no permiten localizar recursos). **Por medio de aplicaciones es posible “des referenciar” las URIs para que simulen un comportamiento de URLs y permitan a las personas navegar a través de los datos.**
- Aportar información relevante a partir del uso de estándares: Los datos almacenados en la Web deben ser estructurados por medio de la utilización de estándares que permitan obtener información relevante al buscar una URI
- Incluir vínculos a otras URIs: Con el objetivo de conectar los datos **conformando una “telaraña” real donde sea posible descubrir todo tipo de información.**

El uso de datos enlazados facilita el uso de URIs HTTP como identificadores globales tanto para los datos como para los términos de vocabularios, el modelo de datos RDF se implementa implícitamente para ser usado a escala global y permite a cualquiera referirse a cualquier cosa (Ríos Hilario et al., 2012, Torre-Bastida et al., 2015). Desde el punto de vista de la búsqueda y recuperación de la información, los usuarios pueden buscar cualquier URI contenida en un grafo RDF sobre la Web para recuperar información adicional. Además, cada tripleta RDF es parte de la Web de los Datos y cada una de ellas puede ser utilizada como un punto de partida para explorar este espacio de datos. Gracias a este modelo de datos es posible establecer vínculos en RDF entre datos de diferentes fuentes. La información proveniente de diferentes fuentes puede ser simplemente mezclada combinando varios conjuntos de tripletas en un mismo grafo RDF permite representar información estructurada bajo esquemas distintos en un mismo grafo, de esta manera se pueden mezclar términos de diferentes vocabularios para describir los datos.

2.8. Referencias bibliográficas

- ALOR-HERNANDEZ, G., SANCHEZ-RAMIREZ, C., CORTES-ROBLES, G., RODRIGUEZ-GONZALEZ, A., GARCIA-ALCARAZ, J. L. & CEDILLO-CAMPOS, M. G. 2014. BROSEMWEB: A brokerage service for e-Procurement using Semantic Web Technologies. *Computers in Industry*, 65, 828-840.
- ANTONIOU, G., FRANCONI, E. & VAN HARMELEN, F. 2005. Introduction to semantic web ontology languages. In: *Reasoning web*, 1-21. Springer Berlin Heidelberg.
- ARPIREZ, J., GÓMEZ-PÉREZ, A., LOZANO, A. & PINTO, H. S. 1998. 2Agent: An ontology-based WWW broker to select ontologies. In: *Workshop on Applications of Ontologies and PSMs*, Brighton, England.
- ASTESIANO, E., BIDOIT, M., KIRCHNER, H., KRIEG-BRÜCKNER, B., MOSESSE, P. D., SANNELLA, D. & TARLECKI, A. 2002. CASL: the common algebraic specification language. *Theoretical Computer Science*, 286, 153-196.
- BAEZA-YATES, R. & RIBEIRO-NETO, B. 1999. *Modern information retrieval*, ACM press New York.
- BAO, J. & MOTIK, B. 2009. rdf: PlainLiteral: A Datatype for RDF Plain Literals. *World Wide Web Consortium (W3C) Recommendation 27 October 2009*. [Online] Available: <https://www.w3.org/2007/OWL/draft/ED-rdf-plain-literal-20090917/all.pdf> [Accessed 8 de abril de 2016]
- BAO, J., KENDALL, E. F., MCGUINNESS, D. L., PATEL-SCHNEIDER, P. F., DING, L. & KHANDELWAL, A. 2009. Owl 2 web ontology language quick reference guide. *World Wide Web Consortium (W3C) Recommendation 27 October 2009*. [Online] Available: <https://www.w3.org/TR/2009/WD-owl2-quick-reference-20090421/all.pdf> [Accessed 8 de abril de 2016]
- BECERRA OCAMPO, O. F. 2013. Clase 3- La semantica y la Linguistica. *Lenguaje y comunicación* [Online]. Available from: http://lenguajeycomprension.blogspot.com/2013_01_01_archive.html [Accessed 5 abril 2016].
- BECHHOFER, S. 2009. OWL: Web ontology language. In: *Encyclopedia of Database Systems* (pp. 2008-2009). Springer US. [Online] Available: http://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_1073 [Accessed 8 de abril de 2016]
- BECKETT, D. 2014. RDF 1.1 N-Triples. *World Wide Web Consortium, Recommendation REC-n-triples-20140225*. [Online] Available: <http://www.w3.org/TR/2014/REC-n-triples-20140225/> [Accessed 7 de abril de 2016]
- BECKETT, D., BERNERS-LEE, T., PRUD'HOMMEAUX, E. & CAROTHERS, G. 2014. RDF 1.1 Turtle. *World Wide Web Consortium*. [Online] Available: <https://www.w3.org/TR/2014/REC-turtle-20140225/> [Accessed 7 de abril de 2016]

- BELTRÁN-GÓMEZ, O. 2013. Implementación de algoritmos para la construcción de proveedores de datos OAI. *educonector. info*, 68.
- BENITO-OSORIO, D., PERIS-ORTIZ, M., ARMENGOT, C. R. & COLINO, A. 2013. Web 5.0: the future of emotional competences in higher education. *Global Business Perspectives*, 1, 274-287.
- BERNERS-LEE, T. 2000. *Semantic Web on XML*. In: *Slides from XML 2000 conference*, Washington DC. [Online] Available: <https://www.w3.org/2000/Talks/1206-xml2k-tbl/slide1-0.html> [Accessed 19 de septiembre de 2015]
- BERNERS-LEE, T., HENDLER, J. & LASSILA, O. 2001. The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284, 34-43.
- BIDOIT, M. & MOSESSE, P. 2004. CASL User Manual, Introduction to Using the Common Algebraic Specification Language. LNCS 2900. Springer-Verlag.
- BIZER, C. & CYGANIAK, R. 2014. RDF 1.1 TriG. *W3C recommendation*, 110. [Online] Available: <http://www.w3.org/TR/trig/> [Accessed 7 de abril de 2016]
- BIZER, C. & SCHULTZ, A. 2009. The Berlin SPARQL Benchmark. *International Journal on Semantic Web and Information Systems*, 5, 1-24.
- BLÁZQUEZ, J., FERNÁNDEZ, M., GARCÍA-PINAR, J. M. & GÓMEZ-PÉREZ, A. 1998. Building ontologies at the knowledge level using the ontology design environment. In: *Proceedings of the Eleventh Workshop on Knowledge Acquisition, Modeling and Management, KAW'98, Banff, Canada*.
- BOWMAN, C. M., DANZIG, P. B., HARDY, D. R., MANBER, U. & SCHWARTZ, M. F. 1994. *Harvest: A scalable, customizable discovery and access system*. DTIC Document. (No. CU-CS-732-94). COLORADO UNIV AT BOULDER DEPT OF COMPUTER SCIENCE.
- BRICKLEY, D. & GUHA, R. V. 2014. RDF Schema 1.1. *W3C recommendation*, 25, 2004-2014. [Online] Available: <http://www.w3.org/TR/rdf-schema> [Accessed 19 de septiembre de 2015]
- BRICKLEY, D. & MILLER, L. 2014. *FOAF Vocabulary Specification 0.99*. Namespace Document 14 January 2014-Paddington Edition. [Online] Available: <http://xmlns.com/foaf/spec/> [Accessed 7 de abril de 2016]
- CALDÓN, E. F., URIBE, G., LÓPEZ, D. M., DE OLIVEIRA, J. P. M. & WIVES, L. K. 2010. Mecanismos de anotación semántica de Contenidos en Plataformas de Redes Sociales. *Cadernos de Informática*, 5, 89-99.
- CAROTHERS, G. 2014. Rdf 1.1 n-quads. *W3C recommendation, Feb*. [Online] Available: <http://www.w3.org/TR/n-quads> [Accessed 19 de septiembre de 2015]

- CARRILLO POZAS, A. 2014. Ontologías: Definición, metodologías y buenas prácticas para su construcción. In: *Conferencia impartida en la Facultad de Ciencias de la Documentación de la UCM* (6 de mayo de 2014), Biblioteca Nacional de España, España. [Online] Available: <http://es.slideshare.net/bne/ontologas-definicion-metodologas-y-buenas-prcticas> [Accessed 13 de enero de 2016]
- CASTELLS, P. 2002b. Aplicación de técnicas de la web semántica. In: *Workshop de investigación en entornos de interacción colectiva (COLINE'02)*. Granada, España. [Online] Available: http://moodle2.unid.edu.mx/dts_cursos_mdI/pos/TI/BE/AM/12/tecnicas_de_la_web_semantica.pdf [Accessed 23 de abril de 2016]
- CASTRO FERNÁNDEZ, R. 2008. *Representación del Conocimiento. Web Semántica*. Universidad Carlos III de Madrid. [Online] Available: <http://www.it.uc3m.es/jvillena/irc/practicas/08-09/05.pdf> [Accessed 19 de septiembre de 2015]
- CAVIERES ABARCA, A., FREDER MENA, S. & RAMÍREZ NOVOA, A. 2010. Tesoros y Web Semántica: Diseño metodológico para estructurar contenidos Web mediante SKOS-Core. *Serie bibliotecología y gestión de información*, 1-64.
- CHIARO, N., GARAT, D., DAMONTE, P. & ACCUOSTO, P. 2005. Generación semiautomática de una ontología para una red de ONG [Ponencia]. In: *XI Congreso Argentino de Ciencias de la Computación*. Argentina.
- CHOWDHURY, G. 2010. *Introduction to modern information retrieval*, Facet publishing.
- CODINA BONILLA, L. & ROVIRA, C. 2006. La web semántica. *Tendencias en documentación digital*. ed. J. Tramullas (Gijón: Trea, 2006), 9-54
- CORCHO, Ó., FERNÁNDEZ, M. & GÓMEZ-PÉREZ, A. 2003. Methodologies, Tools and Languages for Building Ontologies: Where is the Meeting Point?., *Data & Knowledge Engineering*, 46.
- CORLOSQUET, S. J. J. 2009. *Bootstrapping the Web of Data with Drupal*. National University of Ireland, Galway.
- CORRENDO, G., SALVADORES, M., MILLARD, I., GLASER, H. & SHADBOLT, N. 2010. Sparql query rewriting for implementing data integration over linked data. In: *EDBT/ICDT Workshop*, New York, NY, USA. ACM, 1 - 11.
- CROFT, W. B. 1987. Approaches to intelligent information retrieval. *Information Processing & Management*, 23, 249-254.
- DACONTA, M. C., OBRST, L. J. & SMITH, K. T. 2003. *The Semantic Web: a guide to the future of XML, Web services, and knowledge management*, John Wiley & Sons.
- DAVIS, M. 2006. Semantic wave 2006: executive guide to the business value of semantic technologies. USA: Semantic Interoperability Community of

- Practice, 2006. 56 p. *SICoP white paper series, module, 2*. [Online]. Available: <http://web-services.gov/SICOPsemwave2006v1.0.doc> [Accessed 5 de septiembre de 2009]
- DEAN, M., SCHREIBER, G., BECHHOFFER, S., VAN HARMELEN, F., GENDLER, J., HORROCKS, I., MCGUINNESS, D. L., PATEL-SCHNEIDER, P. F. & STEIN, L. A. 2004. OWL web ontology language reference. *W3C recommendation February*, 10. [Online] Available: <http://www.w3.org/TR/owl-ref/> [Accessed 11 de mayo de 2016]
- DING, Y., MITCHELL, S., CORSON-RIKERT, J., LOWE, B. & HE, B. 2011. *The VIVO Ontology: Enabling Networking of Scientists*. [Online] Available: <https://pdfs.semanticscholar.org/4047/750dd7b708b0036ff27b0b31c78faf4ec873.pdf> [Accessed 27 de enero de 2016]
- DING, Y., YAN, E., GHAZINEJAD, A. & JIA, H. 2013. Extending the VIVO ontology to iSchools: Enabling networking of information scientists. *iConference 2013 Proceedings*.
- DOERR, M., GRADMANN, S., HENNICKE, S., ISAAC, A., MEGHINI, C. & VAN DE SOMPEL, H. 2010. El Modelo de Datos de Europeana (EDM). In: *Meeting: 149. Information Technology, Cataloguing, Classification and Indexing with Knowledge Management*.
- DUCHARME, B. 2013. *Learning Sparql : Queryng and Updating with SPARQL 1.1*, Beijing, O ´Reilly.
- EUROPEANA, F. 2016. *Europeana Collections (version beta)* [Online]. Netherlands. Available: <http://www.europeana.eu/portal/> [Accessed 4 de abril 2016].
- EVANS, D. 2011. *Internet de las cosas. Cómo la próxima evolución de Internet lo cambia todo*. Cisco Internet Business Solutions Group (IBSG) [Online] Available: <http://www.cisco.com/web/LA/soluciones/executive/assets/pdf/internet-of-things-iot-ibsg.pdf> [Accessed 4 de mayo 2015]
- FEATHER, J. & STURGES, P. 1997. International encyclopedia of library and information science. London: Routledge.
- GÓMEZ-PÉREZ, A., FERNÁNDEZ, M. & VICENTE, A. D. 1996. Towards a method to conceptualize domain ontologies. In: "*12th European Conference on Artificial Intelligence (ECAI'96)*", Budapest, Rumanía
- GANDON, F. & SCHREIBER, G. 2014. Rdf 1.1 xml syntax. *W3C recommendation 25* February 2014. World Wide Web Consortium. [Online] Available: <http://www.w3.org/TR/rdf-syntax-grammar> [Accessed 19 de septiembre de 2015]
- GARCÍA JIMÉNEZ, A. 2004. Instrumentos de representación del conocimiento: tesauros versus ontologías. *Anales de documentación*, 7, 79-95.

- GARCÍA RUIZ, R. 2008. *TFC: XML y Web Semántica*. [Online] Available: <https://core.ac.uk/download/pdf/9627147.pdf?repositoryId=282> [Accessed 4 de mayo 2015]
- GARCÍA-GARCÍA, A. 2014. *Integración de contenidos semánticos en un portal web de científicos y humanistas valencianos: Vestigium*. Tesis doctoral, Universidad Católica De Valencia San Vicente Mártir.
- GARCIA-MARCO, F. J. 2007. Ontologies and knowledge organization: challenges and opportunities for information professionals. *Profesional De La Informacion*, 16, 541-550.
- GARZÓN, L. 2015. Evolución de la Web. *Tecnología Informática 2015* [Online]. Available: <http://tecnologiainformatica11103.blogspot.com/> [Accessed 19 de marzo 2015].
- GIARETTA, P. & GUARINO, N. 1995. Ontologies and knowledge bases towards a terminological clarification. *Towards very large knowledge bases: knowledge building & knowledge sharing*, 25, 32.
- GOLBREICH, C., WALLACE, E. K. & PATEL-SCHNEIDER, P. 2009. OWL 2 Web Ontology Language: new features and rationale. *W3C working draft, W3C (June 2009)* [Online]. Available: <http://www.w3.org/TR/2009/WD-owl2-new-features-20090611> [Accessed 11 de mayo de 2016].
- GÓMEZ-DUEÑAS, L.-F. 2006. La Iniciativa de Archivos Abiertos (OAI): Un nuevo paradigma en la comunicación científica y el intercambio de información. *Revista Códice*, 2, 21-48.
- GÓMEZ-PÉREZ, A., ANGELE, J., LOPEZ, M. F., CHRISTOPHIDES, V., STUTT, A. & SURE, Y. 2002. *A survey on ontology tools*. [Online]. Available: http://people.aifb.kit.edu/ysu/publications/OntoWeb_Del_1-3.pdf [Accessed 14 de enero de 2009]
- GÓMEZ-PÉREZ, A., FERNÁNDEZ, M. & VICENTE, A. D. 1996. Towards a method to conceptualize domain ontologies. In: "*12th European Conference on Artificial Intelligence (ECAI'96)*", Budapest, Rumanía
- GRANT, J. & BECKETT, D. 2004. RDF test cases. *W3C recommendation*, 10. [Online]. Available: <http://www.w3.org/TR/rdf-testcases/> [Accessed 19 de septiembre de 2015]
- GRAU, B. C., HORROCKS, I., PARSIA, B., RUTTENBERG, A. & SCHNEIDER, M. 2008. *OWL 2 Web Ontology Language: Mapping to RDF Graphs*. [Online]. Available: <https://www.w3.org/TR/2008/WD-owl2-mapping-to-rdf-20081202/all.pdf> [Accessed 19 de septiembre de 2015]
- GRAVANO, L., CHANG, C.-C. K., GARCÍA-MOLINA, H. & PAEPCKE, A. 1997. STARTS: Stanford proposal for Internet meta-searching. *ACM SIGMOD Record*, 26(2), 207-218.
- GRUBER, T. 1993a. *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*. Technical Report KSL- 93-04. CA: Knowledge Systems Laboratory Stanford University.

- GRUBER, T. 1993b. A translation approach to portable ontologies. *Knowledge Acquisition*, 5, 199-220.
- GRÜNINGER, M. & FOX, M. S. 1995. Methodology for the Design and Evaluation of Ontologies. In: *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing*, IJCAI-95, Montreal.
- GUERRERO BOTE, V. P. & LOZANO-TELLO, A. 1999. Vínculos entre las Ontologías y la Biblioteconomía y Documentación. In: La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información. *Actas del IV Congreso ISKO-España EOCONSID'99*, 22-24 de abril de 1999, Granada, Universidad de Granada, 25-31.
- GUZMÁN LUNA, J. A., LÓPEZ BONILLA, M. & DURLEY TORRES, I. 2012. Metodologías y métodos para la construcción de ontologías. *Scientia et Technica*, 2, 133-140.
- HASLHOFER, B. & ISAAC, A. 2011. The Europeana Linked Open Data Pilot. In: *International Conference on Dublin Core and Metadata Applications*. The Hague, The Netherlands 94 - 104.
- HAYES, P. & MCBRIDE, B. 2004. RDF semantics. *W3C recommendation*, 10. [Online]. Available: http://www.open-innovation.in/download/dot/Publication/sub_HaMc2003a.pdf [Accessed 19 de septiembre de 2015]
- HAYES, P. J. & PATEL-SCHNEIDER, P. F. 2014. RDF 1.1 Semantics. *W3C recommendation*, [Online]. Available: <http://www.w3.org/TR/rdf11-mt> [Accessed 19 de septiembre de 2015]
- HEFLIN, J. 2007. An Introduction to the OWL Web Ontology Language. Pensilvania: Lehigh University. National Science Foundation (NSF).
- HIDALGO DELGADO, Y. & RODRÍGUEZ PUENTE, R. 2013. La web semántica: una breve revisión. *Revista Cubana de Ciencias Informáticas*, 7, 76-85.
- HORROCKS, I., FENSEL, D., BROEKSTRA, J., DECKER, S., ERDMANN, M., GOBLE, C., VAN HARMELEN, F., KLEIN, M., STAAB, S. & STUDER, R. 2000. The Ontology Interchange Language OIL. tech. report, Free Univ. of Amsterdam.
- HORROCKS, I., KRÖTZSCH, M. & PARSIA, B. 2009. OWL 2 Web Ontology Language Conformance. *W3C recommendation*. [Online]. Available: <http://www.w3.org/TR/2009/REC-owl2-conformance-20091027/> [Accessed 27 de agosto de 2013]
- ISAAC, A. & SUMMERS, E. 2009. SKOS Simple Knowledge Organization System Primer. *W3C Working Group Note*. (2011-11-03). [Online]. Available: <http://dcevents.dublincore.org/index.php/IntConf/dc-2011/paper/view/69/36> [Accessed 4 de abril 2016]

- JONES, D. M., BENCH-CAPONAND, T.J.M. & VISSER, P. R. S. 1998. Methodologies for Ontology Development. *IT and KNOWS: Conference of the 15th FIP World Computer Congress*.
- KAMBIL, A. 2008. What is your Web 5.0 strategy? *Journal of business strategy*, 29, 56-58.
- KARP, P. D., CHAUDHRI, V. K. & THOMERE, J. 1999. XOL: An XML-based ontology exchange language. [Online]. Available: <http://xml.coverpages.org/xol-03.html> [Accessed 11 de febrero 2016]
- KLYNE, G. & CARROLL, J. J. 2004. Resource description framework (RDF): Concepts and abstract syntax. *W3C recommendation*, 2004. World Wide Web Consortium. [Online]. Available: <http://w3c.org/TR/rdf-concepts> [Accessed 15 de septiembre de 2015].
- KLYNE, G., CARROLL, J. J. & MCBRIDE, B. 2014. RDF 1.1 concepts and abstract syntax. *W3C recommendation*, 25. [Online]. Available: <http://travesia.mcu.es/portalnb/jspui/bitstream/10421/2427/1/RDF%201.pdf> [Accessed 15 de septiembre de 2015].
- KRSULOVIC-MORALES, E. & GUTIÉRREZ, C. 2002. *Propuesta para la Creacion de una Ontologia sobre Departamentos Universitarios de Computacion en Chile*. Informe de investigación, Universidad de Chile.
- KUNA, H. D., REY, M., PODKOWA, L., MARTINI, E. & SOLONEZEN, L. 2014. Expansión de consultas basada en ontologías para un sistema de recuperación de información. In: *XVI Workshop de Investigadores en Ciencias de la Computación*. (Universidad Nacional de Tierra del Fuego, Argentina). [Online]. Available: http://sedici.unlp.edu.ar/bitstream/handle/10915/41513/Documento_completo.pdf?sequence=1 [Accessed 24 de septiembre de 2015].
- LAMARCA, M. J. 2006. *Hipertexto: El nuevo concepto de documento en la cultura de la imagen*. Tesis Doctoral, Universidad Complutense de Madrid.
- LIENDL, M., INNOCENTE, V., CASE, M., VAN LINGEN, F., DE ROECK, A., ARCE, P., FÜRTJES, A. & TODOROV, T. 2001. The Role of XML in the CMS Detector Description Database. CERN-CMS-CR-2001-008.
- LONGLEY, D., KELLOGG, G., LANTHALER, M. & SPORNY, M. 2015. JSON-LD 1.0 Processing Algorithms and API. *World Wide Web Consortium*. [Online]. Available: <http://www.w3.org/TR/json-ld-api/>. [Accessed 18 de febrero de 2016].
- LÓPEZ, M. F., GÓMEZ-PÉREZ, A., SIERRA, J. P. & SIERRA, A. P. 1999. Building a chemical ontology using methontology and the ontology design environment. *Ieee Intelligent Systems*, 37-46.
- LORENTE, M. 2005. Ontología sobre economía y recuperación de información= Ontology for economics and Information Retrieval. *Hipertext. Net*, 3.

- [Online]. Available: <http://www.hipertext.net> [Accessed 23 de abril de 2016].
- LOZANO-TELLO, A., GOMEZ-PEREZ, A. & SOSA, E. 2003. Selection of ontologies for the semantic Web. *In*: LOVELLE, J. M. C., RODRIGUEZ, B. M. G., AGUILAR, L. J., GAYO, J. E. L. & RUIZ, M. D. P. (eds.) *Web Engineering, Proceedings*. Berlin: Springer-Verlag Berlin.
- LUKE, S. & HEFLIN, J. 2000. SHOE 1.01. Proposed specification. *Shoe Project*.
- MANOLA, F., MILLER, E. & MCBRIDE, B. 2004. RDF primer. *W3C recommendation*, 10 (1-107), 6. [Online]. Available: <http://www.uazuay.edu.ec/bibliotecas/mbaTI/pdf/RDF%20Primer.pdf> [Accessed 15 de septiembre de 2015]
- MÁRQUEZ SOLÍS, S. M. 2007. *La Web Semántica*, Chile. [Online] Available: https://books.google.com/cu/books?hl=es&lr=&id=afuncWknStoC&oi=fnd&pg=PA55&dq=SOL%C3%8DS,+S.+M.+2007.+La+Web+Sem%C3%A1ntica,+Lulu.+com.&ots=L6uh1YxKPo&sig=w_wdMGZmXps7bKpHS6EXZakbYI&redir_esc=y#v=onepage&q&f=false [Accessed 7 de mayo de 2015]
- MCGUINNESS, D. L. & VAN HARMELEN, F. 2004. OWL web ontology language overview. *W3C recommendation*, 10(10), 2004. [Online]. Available: http://s3.amazonaws.com/academia.edu.documents/30759881/5.3-B1.pdf?AWSAccessKeyId=AKIAJ56TOJRTWSMTNPEA&Expires=1478816773&Signature=2etuvKxMj2kCFha3ei9d%2FB2M%2B8g%3D&response-content-disposition=inline%3B%20filename%3DOWL_web_ontology_language_overview.pdf [Accessed 17 de febrero de 2016]
- MÉNDEZ RODRÍGUEZ, E. M. 1999. RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio. *In*: *Les biblioteques i els centres de documentació al segle XXI: peça clau de la societat de la informació*, 487-498. [Online]. Available: <http://www.cobdc.org/jornades/7JCD/1.pdf> [Accessed 3 de febrero de 2016]
- MILES, A. & BECHHOFFER, S. 2009. SKOS simple knowledge organization system reference. *W3C recommendation*, 18, [Online]. Available: <http://www.w3.org/TR/skos-reference/> [Accessed 26 de febrero de 2016]
- MOTIK, B., GRAU, B. C., HORROCKS, I., WU, Z., FOKOUE, A. & LUTZ, C. 2009a. Owl 2 web ontology language: Profiles. *W3C recommendation*, 27, 61. [Online]. Available: <https://www.w3.org/2007/OWL/draft/ED-owl2-profiles-20090420/all.pdf> [Accessed 17 de febrero de 2016]
- MOTIK, B., PARSIA, B. & PATEL-SCHNEIDER, P. F. 2009b. OWL 2 web ontology language XML serialization. *W3C recommendation, W3C-World Wide Web Consortium*. [Online]. Available: <https://www.w3.org/2007/OWL/draft/ED-owl2-xml-serialization-20090420/all.pdf> [Accessed 17 de febrero de 2016]

- MOTIK, B., PATEL-SCHNEIDER, P. F. & GRAU, B. C. 2009c. Owl 2 web ontology language direct semantics. *W3C recommendation*, 27. [Online]. Available: <https://www.w3.org/2007/OWL/draft/ED-owl2-direct-semantics-20090914/all.pdf> [Accessed 17 de febrero de 2016]
- MOTIK, B., PATEL-SCHNEIDER, P. F., PARSIA, B., BOCK, C., FOKOUE, A., HAASE, P., HOEKSTRA, R., HORROCKS, I., RUTTENBERG, A. & SATTLER, U. 2009d. OWL 2 web ontology language: Structural specification and functional-style syntax. *W3C recommendation*, 27, 159. [Online]. Available: <http://www.academia.edu/download/42185666/all.pdf> [Accessed 17 de febrero de 2016]
- MUÑOZ, Y. 2015. Evolución de la web y características, Universidad Técnica del Norte. [Online]. Available: <http://es.slideshare.net/AnabelYajaira/evolucion-de-la-web-y-caracteristicas> [Accessed 22 de julio de 2015]
- MYER, T. 2003. *Build an XML-Based Content Management System with PHP* [Online]. Available: <http://www.sitepoint.com/management-system-php/> [Accessed 19 de octubre 2015].
- NAVARRO GALINDO, J. L. 2012. *FLERSA: un sistema semántico de gestión de contenido web (S-CMS)*. Tesis doctoral, Universidad de Granada.
- NECHES, R., FIKES, R. E., FININ, T., GRUBER, T., PATIL, R., SENATOR, T. & SWARTOUT, W. R. 1991. Enabling technology for knowledge sharing. *Ai Magazine*, 12, 36.
- NOY, N. F. & MCGUINNESS, D. L. 2001. Ontology development 101: A guide to creating your first ontology. Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA.
- NOY, N. F. & MCGUINNESS, D. L. 2005. Desarrollo de Ontologías-101: guía para crear tu primera ontología. *traducido del inglés por: E. Antezana* [Online] Available: http://protege.stanford.edu/publications/ontology_development/ontology101-es.pdf [Accessed 13 de enero de 2015]
- OJEDA, H. 2014. Mapa Conceptual Evolución de la Web. *Diplomado Componente Docente en Educación Interactiva a Distancia/ Módulo III: Estrategias, Herramientas y Recursos Educativos*. Barquisimeto: Universidad Fermín Toro.
- PAEPCKE, A., BRANDRIFF, R., JANEY, G., LARSON, R., LUDAESCHER, B., MELNIK, S. & RAGHAVAN, S. 2000. Search middleware and the simple digital library interoperability protocol. *DLIB Magazine.*, 6.
- PARSIA, B. & SIRIN, E. 2004. Pellet: An owl dl reasoner. In: *Third International Semantic Web* [Conference-Poster], Hiroshima, Japan.
- PASTOR SÁNCHEZ, J. & DÍAZ ORTUÑO, P. 2008. SPARQL Lenguaje de consulta para RDF. *Retrieved Abril, 5, 2013*. From Recomendación del W3C de 15

- de enero de 2008. [Online] Available: <http://skos.um.es/TR/rdf-sparql-query/> [Accessed 3 de abril de 2016]
- PASTOR SÁNCHEZ, J. A. 2011. *Tecnologías de la web semántica*, Editorial UOC.
- PASTOR SÁNCHEZ, J. A., MARTINEZ-MENDEZ, F. J. & RODRIGUEZ-MUNOZ, J. V. 2012. SKOS application for interoperability of controlled vocabularies in the field of linked open data. *Profesional De La Informacion*, 21, 245-253.
- PASTOR SÁNCHEZ, J. A., ORDUÑA MALEA, E. & SAORÍN, T. 2013. Mercado semántico automático en gestores de contenidos: integración y cuantificación. *El profesional de la información*, 22, 381-391.
- PEDRAZA JIMÉNEZ, R., CODINA BONILLA, L. & ROVIRA, C. 2007. Web semántica y ontologías en el procesamiento de la información documental. *El profesional de la información*, 16, 569-578.
- PEDRAZA-JIMÉNEZ, R., CODINA, L. & ROVIRA, C. 2007. Semantic web and ontologies in document information processing. *Profesional De La Informacion*, 16, 569-578.
- POPOV, B., KIRYAKOV, A., OGNANOFF, D., MANOV, D. & KIRILOV, A. 2004. KIM—a semantic platform for information extraction and retrieval. *Natural Language Engineering*, 10, 375-392.
- RÍOS HILARIO, A. B., MARTÍN CAMPO, D. & FERRERAS FERNÁNDEZ, T. 2012. Linked data y linked open data: su implantación en una biblioteca digital. El caso de Europeana. *El Profesional de la Información*, 21 (3), 292-297.
- RODRÍGUEZ PEROJO, K. & RONDA LEÓN, R. 2005. Web semántica: un nuevo enfoque para la organización y recuperación de información en el web. *Acimed*, 13(6). Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352005000600003&lng=es&tlng=es [Accessed 7 de mayo de 2015]
- SALTON, G. & MCGILL, M. J. 1983. *Introduction to modern information retrieval*, New York.
- SÁNCHEZ, J. A. P. 2011. *Tecnologías de la web semántica*, Editorial UOC.
- SAVOY, J. & PICARD, J. 2001. Retrieval effectiveness on the web. *Information Processing & Management*, 37, 543-569.
- SCHNEIDER, M., CARROLL, J., HERMAN, I. & PATEL-SCHNEIDER, P. F. 2009. OWL 2 Web Ontology Language: RDF-Based Semantics. *W3C recommendation (October 27 2009)*. [Online]. Available: <https://www.w3.org/2007/OWL/draft/ED-owl2-rdf-based-semantics-20090921/all.pdf> [Accessed 17 de febrero de 2016]
- SENSO RUIZ, J. A. 2003. Herramientas para trabajar con rdf. *El profesional de la información*, 12, 132-139.

- SMITH, B., ASHBURNER, M., ROSSE, C., BARD, J., BUG, W., CEUSTERS, W., GOLDBERG, L. J., EILBECK, K., IRELAND, A. & MUNGALL, C. J. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25, 1251-1255.
- SOLÍS, S. M. 2007. *La Web Semántica*, Lulu. com.
- SOWA, J. F. 2004. Common logic controlled english. Technical report, 2004. Draft, 24 February 2004[Online] Available: <http://www.jfsowa.com/clce/specs.htm> [Accessed 24 de mayo de 2016]
- SPEICHER, S., ARWE, J. & MALHOTRA, A. 2014. Linked data platform 1.0. *Proposed Recommendation, W3C*. [Online]. Available: <http://www.w3.org/TR/2012/WD-ldp-20121025/> [Accessed 5 de Noviembre de 2015].
- SPIVACK, N. 2007. Making sense of the semantic web. [Online]. Available: http://novaspivack.typepad.com/nova_spivacks_weblog/files/nova_spivack_semantic_web_talk.ppt [Accessed 7 de Noviembre de 2015].
- SPORNY, M., LONGLEY, D., KELLOGG, G., LANTHALER, M. & LINDSTRÖM, N. 2014. JSON-LD 1.0. *W3C recommendation (January 16, 2014)*. [Online]. Available: <http://www.w3.org/TR/json-ld/> [Accessed 18 de febrero de 2016].
- STUDER, R., BENJAMINS, R. & FENSEL, D. 1998. Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering*, 25, 161-197.
- SWARTOUT, B., PATIL, R., KNIGHT, K. & RUSS, T. 1996. Toward distributed use of large-scale ontologies. In: *Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems*, 138-148. Banff.
- TABARES, L. & LEIVA MEDEROS, A. A. 2016. *Consideraciones sobre las estructuras de datos formateadas para las consultas en web semántica*. Tesis de maestría, Universidad de las Ciencias Informáticas, La Habana.
- TALLARICO, M. 2003. *Uso de ontologías en tareas de recupero de información*. Tesina de Licenciatura. [Online] Available: <http://usuarios.fceia.unr.edu.ar/~tallaric/tesina/tesina.pdf> [Accessed 28 de mayo de 2016]
- TORRE-BASTIDA, A. I., GONZALEZ-RODRIGUEZ, M. & VILLAR-RODRIGUEZ, E. 2015. Linked open data (LOD) and its implementation in libraries: Initiatives and technologies. *Profesional De La Informacion*, 24, 113-120.
- TRAMULLAS, J. 2011. Los Sistemas de Gestión de Contenidos como pieza fundamental en el despliegue de la Web Semántica. [Online] Available: <http://tramullas.com/los-sistemas-de-gestion-de-contenidos-como-pieza-fundamental-en-el-despliegue-de-la-web-semantica/> [Accessed 24 de febrero de 2014]
- USCHOLD, M. & KING, M. 1995. *Towards a methodology for building ontologies* [Online]. Available:

- <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.480.1214&rep=rep1&type=pdf> [Accessed 11 de enero de 2016].
- VALENCIA CASTILLO, E. 2008. Recuperación y organización de la información a través de RDF usando SPARQL. Universidad Pontificia de Salamanca–Campus de Madrid. [Online]. Available: <http://jena.apache.org/tutorials/sparql.html> [Accessed 15 de septiembre de 2015].
- VAN HEIJST, G., SCHREIBER, A. T. & WIELINGA, B. J. 1997. Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, 46, 183-292.
- W3C 2010. Guía Breve de Linked Data. [SI]: El Consorcio.
- W3C 2010. Guía Breve de Linked Data. [SI]: El Consorcio. [Online] Available: <http://www.w3c.es/divulgacion/guiasbreves/LinkedData> [Accessed 14 de abril de 2016]
- W3C 2012a. *OWL 2 web ontology language document overview*. [Online] Available: <https://www.w3.org/TR/owl2-overview/> [Accessed 8 de abril de 2016]
- W3C. 1999. *Resource Description Framework (RDF): Modelo y Sintaxis Especificación* [Online]. Available: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222> [Accessed 28 de septiembre de 2015].
- W3C. 2012b. *SKOS Simple Knowledge Organization System* [Online]. Available: <http://www.w3.org/2004/02/skos/> [Accessed 10 de diciembre 2015].
- W3C. 2012b. *SKOS Simple Knowledge Organization System* [Online]. W3C. Available: <http://www.w3.org/2004/02/skos/> [Accessed 10. diciembre 2015].
- W3C. 2013. *W3C Semantic Web Activity* [Online]. Available: <https://www.w3.org/2001/sw/> [Accessed 10 de febrero 2016].
- W3C. 2015. *W3Cwww.w3.org* [Online]. Available: <http://www.w3.org> [Accessed 15 de octubre de 2015].
- WALSH, N. 1998. What is XML. *XML. commune*. [Online] Available: http://www.ce.unipr.it/people/bianchi/Teaching/IntelligenzaArtificiale/rdf_pl/XML-RDF/xmlguide1.html [Accessed 8 de abril de 2016]
- YU, L. 2011. *A developer's guide to the semantic Web*. *Springer Science & Business Media*. [Online] Available: https://books.google.com/cu/books?hl=es&lr=&id=G0toUPABjJEC&oi=fnd&pg=PR7&dq=YU,+L.+2011.+A+developer%E2%80%99s+guide+to+the+semantic+Web,+Springer+Science+%26+Business+Media.+&ots=hyHeGxy2Xz&sig=ccBpEow9dApruMu10qOgh2zUvz8&redir_esc=y#v=onepage&q=YU%2C%20L.%202011.%20A%20developer%E2%80%99s%20Oguide%20to%20the%20semantic%20Web%2C%20Springer%20Scienc

[e%20%26%20Business%20Media.&f=false](#) [Accessed 17 de marzo de 2016]

- ZENG, M. & ZUMMER, M. 2009. Introducción de FRSAD y correspondencia con SKOS y otros modelos. In: *World Library and Information Congress, IFLA General Conference and Council*, 75th, Milan. The Hague: IFLA. [Online]. Available: <http://www.ifla.org/files/hq/papers/ifla45/200-zeng-es.pdf> [Accessed 30 de septiembre de 2010].
- ZENG, M. L. & CHAN, L. 2004. Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology*, 55.

CAPÍTULO III
LOS CMS SEMÁNTICOS EN EL CONTEXTO DE
LAS CIENCIAS DE LA INFORMACIÓN:
RELACIONES Y APORTES BIDIRECCIONALES

CAPÍTULO 3: Los CMS semánticos en el contexto de las Ciencias de la Información: relaciones y aportes bidireccionales

En el presente capítulo se aborda desde la teoría el término de S-CMS, su evolución, estructura y modelación. Se presentan otros aspectos que colateralmente inciden en su implementación y desarrollo como: estándares, la anotación semántica, métodos de marcado semántico, datos enlazados, entre otros.

3.1. CMS semánticos (S-CMS): análisis terminológico

Junto con la evolución de las filosofías de la ciencia, han crecido las propuestas para el perfeccionamiento de la actividad informativa en pos del desarrollo cognitivo de la sociedad. Así, en los espacios web que habitualmente desechaban la polisemia y la sinonimia en busca de la exactitud y cientificidad, hoy se intentan acomodar a la diversidad lingüística y sus múltiples significados.

Por otra parte, tradicionalmente la semántica documental desde la CI se ha caracterizado por el uso de modelo de metadatos (dirigidos a identificar el objeto documental) y los lenguajes documentales (o de indización para la representación formalizada de los contenido del documento) como herramientas de la ORI (Luna et al., 2006).

Los CMS en su concepción⁵² inicial y, actualmente en su mayoría, no contemplan la gestión de información semántica como uno de los principios de funcionamiento. Al pasar de los años, con la introducción de la visión de trabajo en el ambiente de la web semántica, se comienzan a incorporar y armonizar nuevos componentes al software que propician el uso de la tecnología semántica en los CMS. En diferentes contextos, aristas y enfoques los autores han abordado el tema, relacionando conceptos provenientes esencialmente de los campos de la CI y la ingeniería de software, dando origen

⁵² El objetivo de un CMS es: (a) permitir la entrada de contenido digital en diversos formatos, (b) permitir la gestión de dicho contenido, y (c) permitir la recuperación posterior de todo esto (Serrano-Cobos, 2007).

a una nueva tipología de CMS vinculado al trabajo de la interpretación semántica de los contenidos, los llamados CMS semánticos.

Los CMS semánticos son sistemas que están emergiendo y aunque “no existe ninguna definición aceptada de momento” (Navarro Galindo, 2012) persiste la tendencia a realizar extrapolaciones que los definen como los sistemas donde se integran las herramientas de los CMS con herramientas semánticas (Martínez Usero, 2004)⁵³.

Bajo esta misma dirección los definen Christ y Nagel (2011) quienes los delimitan como aquellos “CMS con la capacidad de interactuar con, la extracción, la gestión, y el almacenamiento de metadatos semánticos sobre el contenido” (Pastor Sánchez, 2011b), características que le son reconocidas de manera consensuada.

A los efectos de la presente investigación se entenderá que **un CMS semántico** (o S-CMS, como se ha tornado común la referencia al tema en la literatura por sus siglas en inglés: semantic content management system) **es un CMS⁵⁴ que se apoya en tecnologías semánticas combinadas para aportar descripciones de contenido explícito de los recursos de manera que sean entendibles por un software.** Básicamente se encamina a hacer posible las ideas paradigmáticas en torno a los procesos ORI⁵⁵ en la web semántica, **al permitir procesos inteligentes de captura y tratamiento de la información.**

⁵³ Citado por NAVARRO GALINDO, J. L. 2012. *FLERSA: un sistema semántico de gestión de contenido web (S-CMS)*, Universidad de Granada.

⁵⁴ Por tanto, para los fines de la presente investigación, se parte del concepto de CMS aportado por la autora en el capítulo 1 cuando asume que: un CMS es una herramienta informática, que funcionan sobre la base de una filosofía de holística, estructurada en módulos que permiten la creación, almacenamiento, actualización, recuperación y visualización de los contenidos, a distintos niveles de gestión y acceso, en función del producto implementado. Es un sistema sinérgico, dinámico e integral, que permite manejar de manera independiente el diseño visual, de la gestión del contenido.

⁵⁵ Organización y Recuperación de la Información

Los S-CMS son el resultado de la búsqueda de una gestión más eficiente para la gestión de información en la web, siendo reflejo de las ideas vanguardistas dentro de la CI en los últimos años.

3.2. Características

Como producto derivado de un software más genérico como lo es el CMS, los S-CMS heredan muchas de sus características pero se han ido adaptando al sistema de trabajo de los nuevos requerimientos y expectativas.

Entre las particularidades comunes se pueden considerar:

- Espacio de trabajo colaborativo y, por lo general, centrado en el usuario como ente fundamental en la asignación y gestión de valores semánticos del contenido (dado por el contexto de creación y utilización del dominio)
- Uso de estándares y especificaciones de representación de datos (XML, RDF, OWL, SKOS, SPARQL) para el manejo y accesibilidad de recursos diversos de manera simultánea y que a su vez permiten compartir, vincular y tributar al enriquecimiento de la información semántica gestionada en el sistema (anotaciones semánticas de los contenidos tanto manuales como automatizadas, ontologías, etc.)
- Buscan la interoperabilidad y el intercambio de información en sí mismos y con otros sistemas. El software no debe generar conflictos al establecer relaciones y/o crear nuevos contenidos semánticos.
- Las connotaciones semánticas y la relación entre los contenidos son dinámicas y evolucionan en correspondencia al comportamiento léxico del dominio que lo genera, por tanto, las herramientas empleadas para materializar dicha características a través del S-CMS se distinguen por su alto grado de flexibilidad y estandarizaciones.

3.3. S-CMS y gestión del contenido

3.3.1. El contenido

Tal vez ha sido la interdisciplinaridad del término “contenido” lo que ha matizado su carácter polisémico y controversial. La imposibilidad de restringirlo a un solo campo de acción, influye en la multiplicidad de definiciones y perspectivas de estudio relacionadas al mismo.

A los efectos de la presente investigación, el “contenido” se expresa en dos dimensiones: tecnológica e informativa, como consecuencia no solo de la yuxtaposición de la CI y la Informática, también como resultado de las transformaciones que han ocurrido en la Sociedad a partir del desarrollo social y tecnológico impulsado, especialmente, por la aparición y socialización de Internet, cuya consecuencia ha sido el enfrentamiento a nuevas problemáticas en la representación de contenidos. Bajo esta realidad, los documentos se combinan en un nuevo entorno de comunicación permeada por la hipertextualidad con diversas características semiológicas y la combinación de texto, sonido e imágenes.

Desde la Informática, el contenido es interpretado como un “sistema de códigos binarios”⁵⁶. Mediante el empleo de este código como base, se crea,

⁵⁶ Codina (2001) con respecto al término “sistema de códigos binarios”, como la forma de interpretación que el ordenador da a la información, plantea: “Decimos que una información es digital cuando está codificada en el formato que puede interpretar un ordenador y suele decirse que consiste en series de ceros y de unos (0, 1). Pero esto, aunque útil, no deja de ser una simplificación. Ningún ordenador puede saber qué es un cero o qué es un uno. Por tanto, cada vez que decimos que procesa series de estos números lo estamos “humanizando”. En realidad solamente registra cambios en señales que pueden adoptar cada vez uno de dos estados posibles” (Codina, Lluís. “Las propiedades de la información digital”. En: *El profesional de la información 2001, diciembre, v. 10, n. 12, pp. 18-25.*).

En esencia, por esta razón autores como Deleuze y Guattari (1995) proponen comprender las formas simbólicas del ciberespacio como metáforas, porque no se confinan al hecho físico de la realización, a la rigidez temporal resultante del registro material y, sobre todo, porque expanden esencias semánticas (Deleuze G, Guattari F. Mil: capitalismo e esquizofrenia. São Paulo: Editora 34,1995. Citado por: BOSCH, M. 2001. Documentos y lenguaje de marcado: conceptos, problemas y tendencias. *El profesional de la información, 10, 4-9.*).

implementa y desarrolla el CMS, lo que hace que habitualmente el significado de los términos no sea comprendido (semánticamente) por el software que lo gestiona. El sistema puede realizar cómodamente diferentes equivalencias pero no inferencias. En este contexto el término (contenido) es generalmente asociado a determinados conceptos como: datos estructurados, metadatos y paquete informático.

Desde la CI, el **concepto de "contenido" ha sido abordado desde múltiples enfoques y disciplinas**. Para esta área del conocimiento puede implicar información, materia, datos, asunto, documento, en dependencia del contexto en que se utilice. Los profesionales de dicho gremio han estudiado los diferentes procesos que condicionan y transforman la percepción semántica del término.

Bajo la dimensión informativa, el "contenido" enunciado en un sitio Web, es reflejo del discurso como un acontecimiento comunicativo y modo de interacción social. El significado y connotación del término (contenido), así como el análisis de sus implicaciones para los procesos de organización y representación de la información, ha estado condicionado a paradigmas y enfoques teóricos que han distinguido las formas de hacer y pensar esta ciencia. Dentro de la influencia de la filosofía de la ciencia, se distinguen en la CI tres paradigmas dominantes (Bosch, 2001):

- **Paradigma Físico:** A partir de su influencia, se fundamentaron estrategias datológicas para la organización, representación y recuperación de las estructuras formales de los documentos, que dan lugar a representaciones y modelos de recuperación léxico-sintácticos.
- **Paradigma Cognitivo:** Se caracterizó por insertar en la CI el estudio de las estructuras mentales individuales de los usuarios como forma de conocer su conducta, para el comienzo de representaciones y dimensiones semánticas del contenido.
- **Paradigma Sociocognitivo:** Llama la atención sobre el reconocimiento del contexto social en las formas de entender y representar la realidad, desde una posición de interacción gnoseológica individuo – grupo social.

El sociocognitismo impulsa el método de análisis de dominio como estrategia para organizar y representar el conocimiento mediante la interpretación del conjunto de relaciones significativas que existen en el contexto documental y extradocumental.

En los años transcurridos posteriores al 2000, en la CI los mecanismos implementados desde los CMS para la representación del contenido y su posterior búsqueda y recuperación, siguen la línea de pensamiento del paradigma cognitivo. Los CMS tradicionales permiten la sindicación de los contenidos, a partir de los RSS para la indización automatizada, pero no reconoce la polisemia o la sinonimia ni las relaciones significativas que dan sentido semántico al contenido para las opciones e inferencias en recuperación de la información. Sobre dichas debilidades trabajan los S-CMS, bajo la visión e influencia del paradigma sociocognitivo en los procesos metarepresentacionales de los contenidos, a partir de anotaciones semánticas y ontologías como principales exponentes de la web semántica.

El contenido con valor semántico, como concepto, es una de las líneas de trabajo fundamental bajo la filosofía de construcción de una web semántica en pos de compartir, integrar y recuperar recursos. Por tanto, desde los S-CMS se intenta optimizar y enriquecer la estructura de la información y añadiendo elementos semánticos que sean entendibles por el software. La nueva generación de formatos está encabezada por XML (Extensible Markup Language) y RDF (Resource Description Framework), los cuales incluirán ontologías- taxonomía de conceptos con atributos y relaciones que proporcionan un vocabulario consensuado para definir redes semánticas de unidades de información interrelacionadas - que especifican las reglas lógicas para que los agentes de software reconozcan y clasifiquen cada concepto (Castells, 2002b).

Las ideas presentadas pueden resumirse a partir del esquema de la figura 3.1:

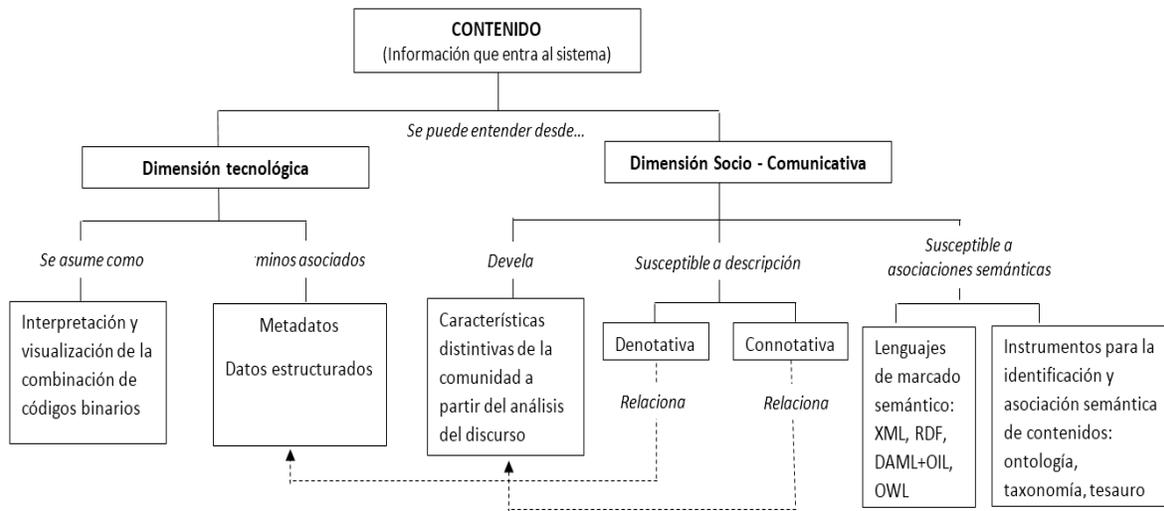


Fig. 3.1. Implicaciones del término "contenido" para el estudio de los CMS

A partir de las ideas analizadas se muestra, una vez más, el carácter multidimensional que matiza el estudio de los S-CMS.

La gestión de información en y con los S-CMS se ha convertido en un verdadero paradigma que ha generado en los últimos años un auge en su estudio y desarrollo.

La incorporación de herramientas de definición de diferentes tipos de contenidos, la aplicación de metadatos o la creación y aplicación de taxonomías, clasificaciones, esquemas jerárquicos y/o asociativos para deducir relaciones entre contenidos, ha permitido un salto cualitativo trascendental en el planteamiento de los CMS, apuntando a la organización de los contenidos como su aspecto clave (Pastor Sánchez and Martínez Méndez, 2009).

3.3.2. Estructura del S-CMS para la gestión de contenido semántico

En varios proyectos se han modelado las características de los S-CMS desde casos y propuestas concretas. En el presente epígrafe se presentan y analizan dos de ellos.

Proyecto IKS, 2009

El proyecto Interactive Knowledge Stack (IKS)⁵⁷, se inicia en 2009 con la construcción de una plataforma de tecnología abierta y flexible para S-CMS. Para ellos la arquitectura de un S-CMS requiere de cuatro capas que se muestran en la Figura 3.2.

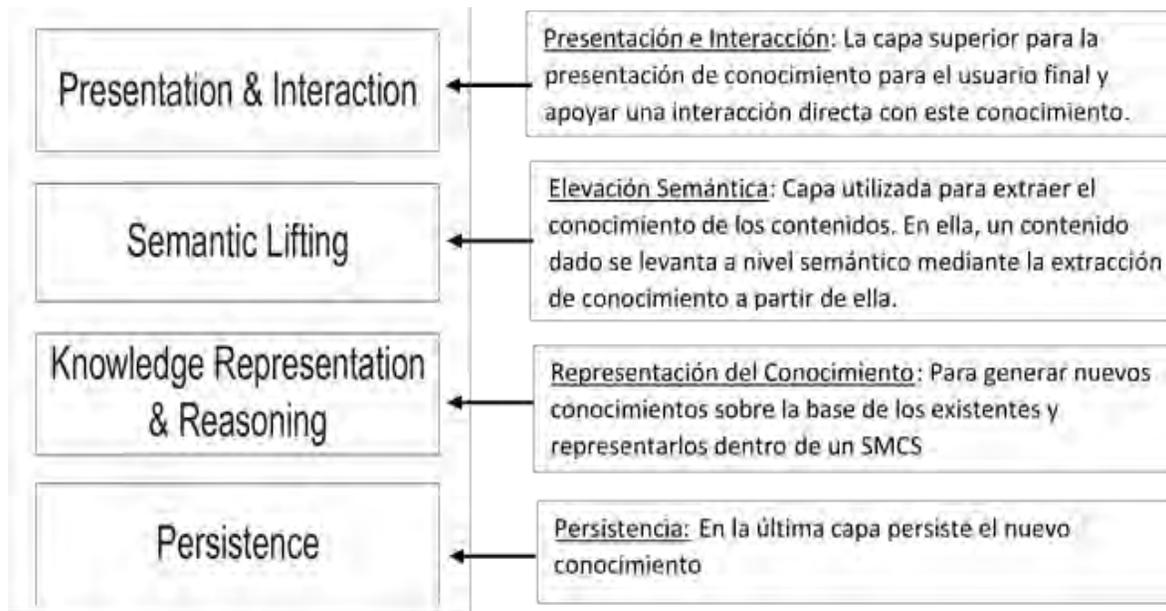


Fig. 3.2. Representación gráfica de la estructura genérica de capas de un S-CMS (CHRIST, 2012)

En un S-CMS el contenido se almacena en un depósito de contenido tradicional. La propuesta de arquitectura de un S-CMS, como se muestra en la Figura 3.3, parte de que cualquier CMS existente, que tiene una arquitectura similar a la representada en la Figura 3.2, se puede ampliar para convertirse en un S-CMS.

Como modelo conceptual es coherente con una de las características de los S-CMS: la reutilización de contenidos y conocimiento, pues propone que para crear un S-CMS de un CMS la columna contenidos tradicionales (Fig. 3.1) se

⁵⁷El proyecto IKS terminó en marzo de 2013. Los principales resultados de IKS están disponibles bajo la **Sección de demostración** (http://www.iks-project.eu/Demos&usg=ALkJrhiyoEjxJ3B-ZMOMsM9Q70_q5As1mA). Para más información sobre los acontecimientos ocurridos desde 2013, puede consultarse **Redlink** (<http://redlink.co/>) o en **SRFG / KMT** (http://www.salzburgresearch.at/en/bereiche/kmt_en/).

debe extender por una segunda columna de conocimientos para los rasgos semánticos en paralelo. Donde al interactuar entre sí ambas columnas se conecten al nivel de acceso al contenido y acceso a los conocimientos.

El esquema (Fig.3.3) muestra como el contenido del S-CMS se almacena en la columna de contenido. Este es transportado desde la columna de contenido a la columna de los conocimientos en la capa de acceso al contenido / Conocimiento. El contenido entregado se puede analizar por los componentes de la columna de conocimiento. El conocimiento obtenido se almacena en la columna de conocimiento. Una vez que la columna de contenido tiene conocimiento adicional con respecto a algunos contenidos puede consultar la columna de contenido. La columna de conocimiento puede buscar contenido nuevo o modificado en la columna de contenido(Christ, 2012).

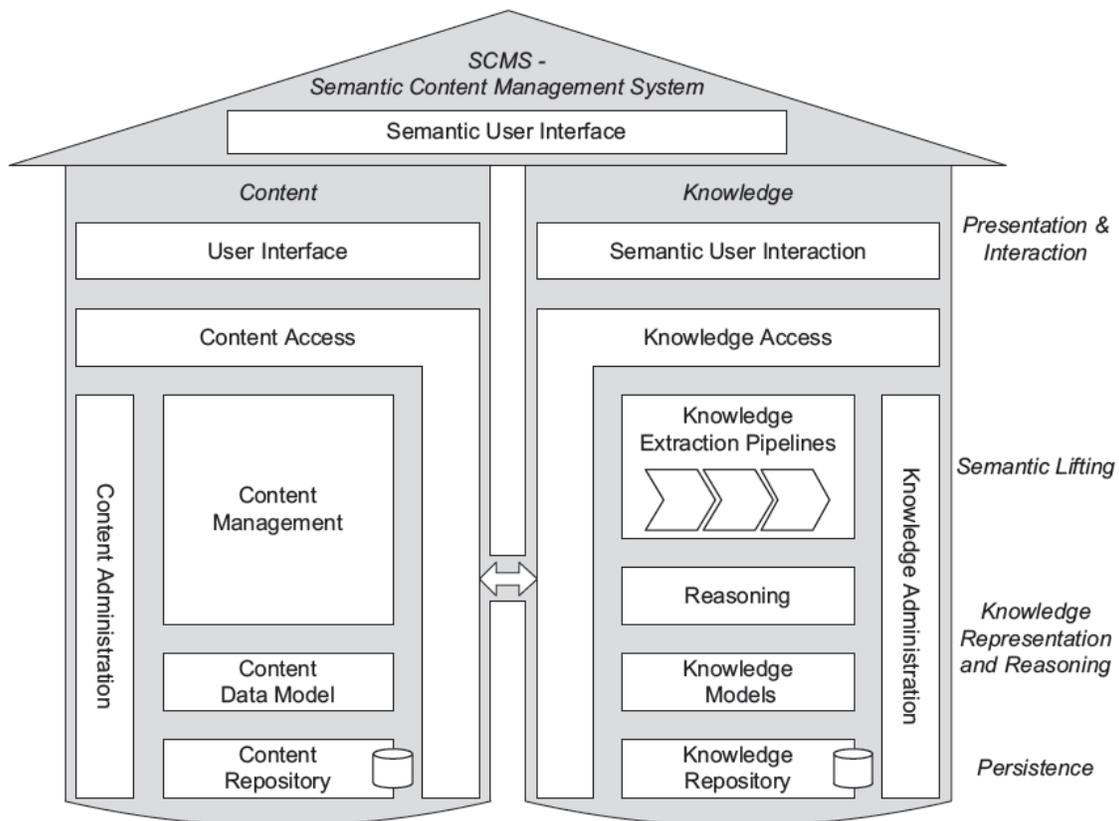


Fig. 3.3. Estructura compleja de capas de un S-CMS (Christ, 2012)

El modelo prevé cuatro capas de actividad (Fig. 3.2 y 3.3):

- **Presentación e Interacción:** En un CMS tradicional, el usuario es capaz de editar y utilizar los contenidos a través de una interfaz de usuario. Cuando se trata de conocimiento en S-CMS se necesita una capa adicional a nivel de interfaz de usuario, llamada **Semantic User Interaction**, que permite al usuario interactuar con el contenido.
- **Elevación Semántica:** Uno de los problemas para los CMS tradicionales es la falta de capacidad para extraer conocimiento en términos de metadatos semánticos del contenido almacenado. Por lo tanto, para un S-CMS se define una capa de Secuencia de Extracción de Conocimiento (**Knowledge Extraction Pipelines**) que encapsulan algoritmos para la extracción de metadatos semánticos. Por lo general, la extracción de conocimiento es un proceso de varias etapas mediante la aplicación de algoritmos conocidos en el campo de la investigación de la extracción de información y recuperación.
- **Representación y Razonamiento:** Después de levantar el contenido a nivel semántico, la información extraída puede utilizarse como plataforma para las técnicas de inferencias en la Capa de Razonamiento. El **razonamiento lógico** es una técnica de inteligencia artificial que utiliza relaciones semánticas para recuperar el conocimiento sobre el contenido que no se conocía explícitamente antes.
- **Persistencia:** El conocimiento se almacena en un repositorio de conocimiento que define la estructura fundamental de datos para el conocimiento. Influenciado por las ideas de la web semántica un triple (relación establecida entre sujeto, el predicado y el objeto) se puede utilizar para expresar cualquier relación entre un sujeto y un objeto. Para hacer de esto una relación semántica uno tiene que definir los modelos de conocimiento en la parte superior del depósito de conocimiento, por ejemplo, para especificar el significado semántico de un cierto predicado.

Proyecto FLERSA, 2012

FLERSA (*Flexible Range Semantic Anotation*) es una herramienta de anotación semántica de contenido web centrado en el usuario, propuesta como resultado de la investigación doctoral de Navarro Galindo (2012). Propone como modelación de estructura de un S-CMS 4 niveles esenciales (Navarro Galindo, 2012):

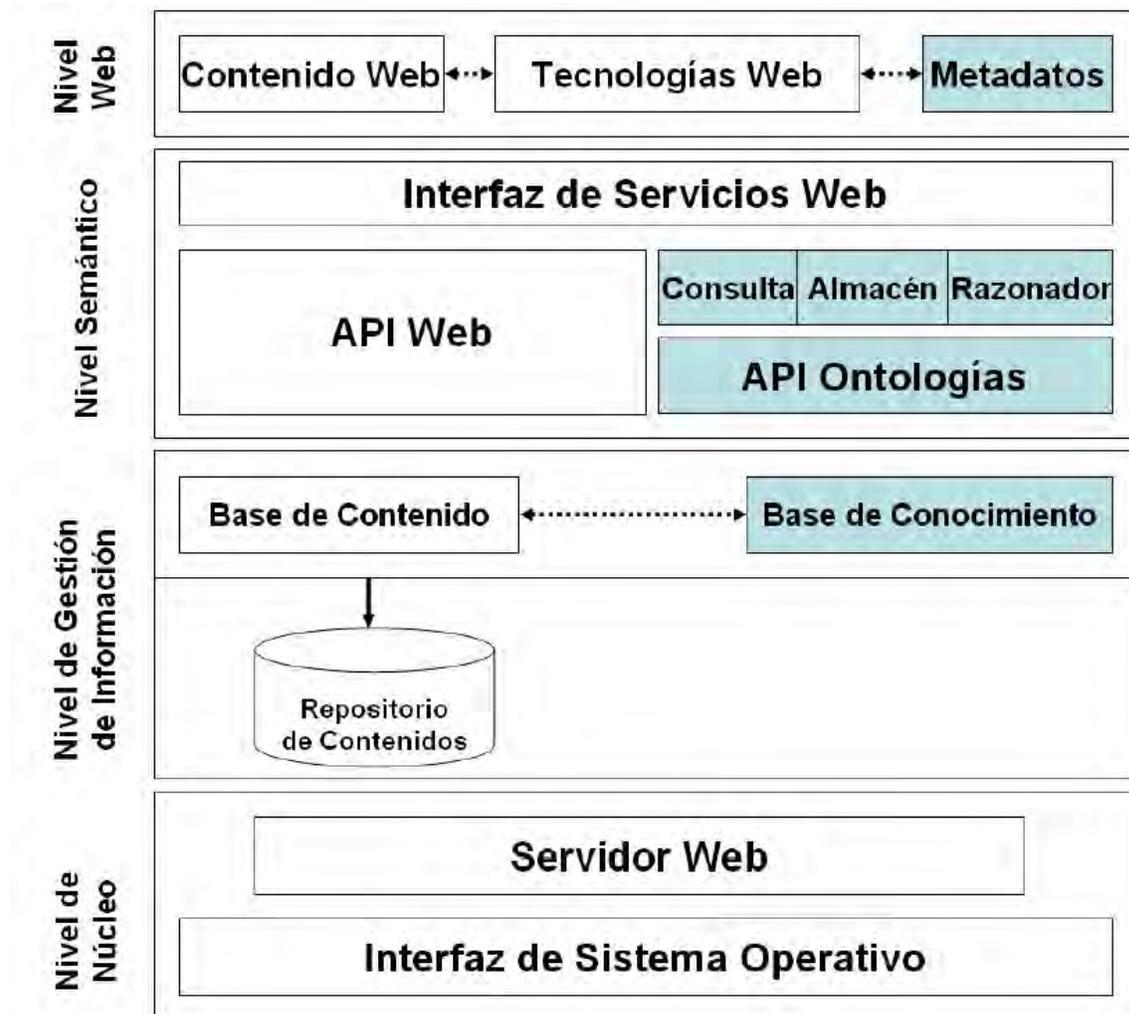


Fig. 3.4. Estructura del S-CMS FLERSA (Navarro Galindo, 2012)

- **Nivel Núcleo:** Aporta la infraestructura software y de comunicaciones necesaria para instalar el resto de los componentes del sistema. Contiene el sistema operativo, los servicios de red y el servidor web.
- **Nivel de Gestión de Información:** Está formado por un componente del sistema llamado *repositorio de contenidos* encargado del almacenamiento

tanto del contenido web (base del contenido). También es el nivel donde se almacena la **base de conocimiento** compuesta por las diversas ontologías con las que trabaja el sistema. Por tanto, se va a necesitar un gestor de base de datos para el alojamiento de la información del repositorio de contenidos.

- **Nivel semántico:** se desarrollan los servicios de aplicación del servidor a través de la interfaz de servicios web. En este nivel confluyen los servicios de contenidos, encargados de trabajar con información dirigida a las personas, con los servicios semánticos, encargados de enriquecer esta información con metadatos dirigidos a las computadoras.
- **Nivel de Interfaz Web:** En este nivel se realiza toda la interacción con el S-CMS. Conviven los contenidos de los documentos web, los metadatos vinculados a estos y las tecnologías web encargadas de modificar en tiempo de ejecución los documentos web para dotarlos de anotaciones semánticas en forma de metadatos, así como de realizar la gestión de mensajes oportuna, mediante el uso de servicios del lado servidor, para aportar la funcionalidad de la herramienta.

Como elementos comunes de las modelaciones teóricas anteriores, se advierte que aparecen en los extremos de representación de la abstracción de modelado del S-CMS:

- En la parte superior la interfaz para el intercambio con el ambiente web y en la base del diagrama la estructura destinada al almacenamiento de la información.
- De igual forma para todos los ejemplos descritos se establece un nivel semántico entre el sistema y su visualización e intercambio con la web como elemento que los distinguen de los tradicionales CMS.

A diferencia de la estructura habitual de un CMS donde el contenido se **almacena en una base de datos "limitada" al propio sistema**, el CMS semántico precisa de un espacio de intercambio e interpretación del contenido a partir de inferencias. Demanda de la interacción con otros sistemas, no de manera arbitraria sino, de forma estructurada, jerárquica, interrelacionando los

términos para que el software perciba la dimensión semántica de los mismos en los procesos de gestión de contenido.

Dada la connotación del significado que para el software cobran los términos, se estaría en presencia de un proceso de gestión más complejo que de contenido, a nuestro entender se gestiona conocimiento.

Ello hace que la modelación conceptual de un CMS semántico tenga un alto grado de subjetividad. La interoperabilidad más que una ventaja de uso de los S-CMS es una necesidad para su funcionamiento, que va a permitir la reutilización de contenidos y relación entre términos y todo tipo de objetos en la red.

3.4. Procesos de organización y recuperación de la información desde los S-CMS

Los procesos de organización y representación de la información (ORI), han sido abordados en la CI como uno de los temas medulares de su desarrollo. El ciclo de vida de la información (CVI), como principio de trabajo, se ha asumido como teoría rectora en disímiles de espacios para el desarrollo de soluciones en diversos contextos de acción de nuestros profesionales.

El desarrollo generado por la aparición y socialización de Internet, trajo consigo nuevas problemáticas, especialmente en la organización y recuperación de los contenidos publicados en la Web. Señalan (Castells, 2002a) **que se trata de “una transformación sociotecnológica” a nivel macro y micro social**, donde las condiciones de generación de conocimiento y procesamiento de la información se han alterado sustancialmente debido a una **revolución tecnológica**. “Los nuevos documentos que circulan en las estructuras reticulares presentan plataformas lógicas que se actualizan mediante mecanismos de inteligencia artificial combinando texto, sonido, imágenes, que se combinan en un nuevo entorno de comunicación con diversas características **semiológicas**” (Leiva Mederos, 2011). “Desde el punto de vista teórico, la reciente gestación de un campo epistémico que examina problemas de la dinámica social de la información, obliga a compartir espacios de reflexión desde la convergencia de disciplinas sociales e informáticas en

virtud de una finalidad: la recuperación oportuna, eficiente, fidedigna e **inteligente**" (Bosch, 2001).

En la CI el estudio y búsqueda de filosofías y enfoques para lograr representaciones lógicas y ergonómicas de la información en sus diferentes manifestaciones, ha sido un foco de atención; sin dejar de reconocer que son temas interdisciplinarios que requieren la imbricación de otras áreas de conocimiento como la Psicología cognitiva, Arquitectura de la información, Lingüística, Informática, ingeniería de software, entre otras.

El tema de la organización y recuperación de los contenidos en el campo de los S-CMS, se expresa en dos dimensiones: filosófica y tecnológica, como resultado del desarrollo de las tecnologías y las implicaciones de su uso en la relación información – tecnología – comunicación en un entorno de red, social y digital.

La dimensión filosófica en los S-CMS, trasciende el uso de la herramienta. Implica un análisis desde la conceptualización de la gestión de contenidos reconociendo al software capaz de asumir e inferir la semántica de los términos en el espacio digital. Apuesta por funcionalidades relacionadas con la creación, relación e inferencias de los contenidos, ya sea mediante la edición directa, la conversión de archivos, o mediante agregación o adquisición (incluye la redifusión o sindicación).

Desde una dimensión tecnológica, los CMS (tradicionales), incorporan un grupo de herramientas que facilitan la entrada, creación y publicación de contenidos en el sistema, en múltiples formatos, añadiendo metadatos a los mismos. Osuna y Cruz (2010), acotan que permiten la recuperación desde distintos ámbitos: "temporal, por categorías, por autor...Terminando con la función de mantenimiento de los contenidos, la actualización de los enlaces, la visualización, todo ello siguiendo unos criterios de calidad y normalización en cuanto a su forma, que configurarán **un sitio permanente y estable**"(Osuna Alarcón and De la Cruz Gómez, 2010). Cuenta con elementos que tributan a la indización de los contenidos de forma automatizada y, como consecuencia lógica, a la organización y recuperación de los contenidos:

- ***Sindicación del contenido RSS***: RSS (Really Simple Syndication) es parte de la familia de formatos XML. De estructura simple comparado con otros estándares.

La sindicación de contenidos ofrece ventajas en el seguimiento de la información nueva y con ella los especialistas y usuarios pueden mantenerse al tanto de los principales resultados publicados ahorrando tiempo y esfuerzo de navegación.

- ***Compatibilidad con XML***: Entre sus objetivos está el lograr una Web semántica donde la información y la forma de representar estén claramente separadas. Así exhibe la posibilidad de trabajar de manera independiente la información y el estilo o formato del documento.

A pesar del desarrollo del XML, sus lenguajes derivados, los esquemas de descripción de recursos de información y de metadatos; varios autores (Walsh, 1998, Tramullas, 2005, Bratsas et al., 2012) han considerado que el desarrollo de las soluciones ofrecidas para la gestión de contenidos y el soporte para la utilización de RDF o de ontologías, no han estado en correspondencia con las demandas informacionales emergentes dada la celeridad y el crecimiento exponencial de la información en la web, dada la actual orientación de las normas.

- ***Buscador o motor de búsqueda***: Herramienta que permite al usuario acceder a la información sobre un tema determinado contenido en el sitio Web, a través de palabras clave introducidas por él. Una de las filosofías de los CMS semánticos es la reutilización de los contenidos, y una manera de ayudar a ellos es disponer de un buscador interno. Facilita que un autor, antes de crear un contenido o un componente, el busque en el repositorio por si puede reutilizarlo, funcionalidad muy útil en lugares distribuidos y con gran cantidad de contenidos.
- ***Metadata***: Herramienta para la recuperación de información sobre objetos Web. Controla la forma en que los sitios Web son indexados por los motores de búsqueda.

Osuna y Cruz definen tres tipos de categorías de metadatos para describir objetos digitales: (Osuna Alarcón and De la Cruz Gómez, 2010)

- Descriptivos: con información que describe el contenido intelectual del objeto, tales como la catalogación de registros MARC, localización o sistemas similares. Normalmente se utiliza para fines bibliográficos y para la búsqueda y recuperación.
- Estructurales: con información que vincula partes del documento a otras para formar unidades lógicas (por ejemplo, la información que relaciona las imágenes de páginas individuales de un libro a los demás componentes del libro).
- Administrativos: con información necesaria para la gestión del objeto o del control de acceso. Esto puede incluir información sobre la forma en que se ha escaneado, el formato de almacenamiento, información sobre licencias y derechos de autor, y toda la información necesaria para la preservación a largo plazo de los objetos digitales.

Aun cuando estas etiquetas sirven para describir la información de los documentos, no garantizan la flexibilidad y la semántica necesaria. El CMS tradicional tiene un conjunto de etiquetas prefijadas y finitas que resultan insuficientes, en muchos casos, para una descripción cabal de determinados recursos de información. "La mayor parte de ellos organizan los contenidos usando un conjunto de metadatos descriptivos (título, autor, fecha) asociados a un campo de texto principal sobre el que se pueden aplicar distintos tipos de formatos, definir enlaces o insertar imágenes" (Pastor Sánchez et al., 2013). Así, los CMS tradicionales y la Web actual se enfrentan a dos retos: definir el significado semántico de los contenidos y facilitar la reutilización de la información de los mismos (Pastor Sánchez, 2013). Esta nueva etapa ya había sido reconocida y declarada por algunos autores en trabajos publicados en las puertas del siglo XXI. Ante el reto de la organización y recuperación de información en la web, los lenguajes de marca estaban llamados a convertirse en el nuevo alfabeto semántico para

los productores y para los gestores de información (Méndez Rodríguez, 1999).

La calidad de un S-CMS depende en gran medida de las posibilidades que ofrezca para el procesamiento del lenguaje natural. Pero ellos por sí mismos son insuficientes. Es necesario vincularlos a ontologías, mecanismos de anotación semántica, estándares, etc. Por tanto se torna necesario el vínculo con otras herramientas y el auxilio de diferentes lenguajes que le permiten desarrollarse y participar en el "ciclo de vida de la gestión de conjuntos de datos a partir del marcado semántico de contenidos" (Pastor Sánchez et al., 2013).

3.4.1. Estándares y lenguajes de marcado más utilizados en los CMS semánticos

La mayoría de la información que se encuentra actualmente en Internet es textual y sin ninguna estructura formal que permita a las máquinas entender y aprovechar dicha información de manera automática y con ella trabaja el CMS tradicional. Sin embargo, en una red social todos los contenidos que están compartiendo son valiosos, por lo tanto es necesario definir mecanismos para dar estructura y enriquecer semánticamente la información existente, para que esta sea aprovechada por otras aplicaciones software (Caldón et al., 2010). Tanto para la realización de técnicas de procesamiento de lenguaje natural, de anotaciones semánticas en sus distintas variantes, o de cualquier tecnología semántica como los S-CMS, requieren de estándares, lenguajes de marcado especializados para lograr representar, identificar las estructuras en los documentos, agrupar y relacionar los conceptos de un dominio de conocimiento, de la manera menos ambigua posible, facilitando su recuperación.

Ante el reto de la organización y recuperación de información digital, los lenguajes de marca se han convertido en un nuevo alfabeto semántico para los productores y para los gestores de información (Méndez Rodríguez, 1999).

En los últimos años se han creado numerosas aplicaciones en pos de mejorar las condiciones representacionales y de recuperación de información en la Web. Los principales medios con los cuales se persiguen los objetivos de la web semántica son: mediante una codificación de páginas en la cual las etiquetas transporten carga semántica (XML) y aportando descripciones (metadatos) de las páginas y sitios web con un formato que sea compatible con la estructura general de la WWW y con diversas categorías de páginas, e interoperable entre distintos sistemas informáticos (RDF). A todo ello se integra un sistema de ontologías que permita especificar conceptos de diversos dominios mediante el uso de un lenguaje basado en lógica simbólica y susceptible a ser interpretado por un ordenador (OWL) (Pedraza Jiménez et al., 2007).

En los sub-epígrafes siguientes se abordarán dichos estándares para el entorno de la Web Semántica bajo una visión general de las características que aporta cada uno, XML, RDF y OWL, tomando en consideración que sobre los mismos se enriquecen las estrategias de trabajo de los S-CMS semánticos.

3.4.1.1. XML: *Extensible Markup Language*⁵⁸

Dadas sus características y potencialidades, XML es la capa sintáctica base de la web semántica. Todas las otras tecnologías que proporcionan características de la web semántica se construirán sobre la base de XML (Daconta et al., 2003). En un CMS contribuye esencialmente como lenguaje de descripción y además es muy útil para la migración de datos (Liendl et al., 2001). Son varios los CMS con aplicaciones semánticas que se implementan con este lenguaje. Entre los más conocidos puede citarse a WordPress, Joomla o Drupal.

3.4.1.2. RDF: *Resource Description Framework*⁵⁹

RDF se puede utilizar en una variedad de áreas dentro del procesamiento de la información desde el CMS y bajo la visión de la CI; por ejemplo: en la

⁵⁸ El epígrafe se trabajó, esencialmente, con los contenidos extraídos de www.w3.org

⁵⁹ El epígrafe se trabajó con los contenidos extraídos de www.w3.org

búsqueda de recursos para proporcionar mejores capacidades de los motores de búsqueda, en la catalogación para describir el contenido y las relaciones de contenido disponible en un sitio en particular Internet, página, o biblioteca digital, por parte de agentes de software inteligentes para facilitar el intercambio de conocimientos y el intercambio, en la clasificación de contenidos, en la descripción de colecciones de páginas que representan una única lógica "documento"; para describir los derechos de propiedad intelectual de las páginas Web; para expresar las preferencias de privacidad de un usuario, así como las políticas de privacidad de un sitio Web (W3C, 1999).

Es la infraestructura que permite la restricción, reutilización e intercambio de metadatos estructurados. Tiene características que facilitan la fusión de datos, incluso si los esquemas subyacentes difieren, y soporta específicamente la evolución de esquemas en el tiempo sin necesidad de todos los consumidores de datos que pueden cambiar.

Por otra parte, RDF Schema es débil en cuanto a principios semánticos, lo que constituyó la causa fundamental para el desarrollo de otros estándares como OWL o SKOS que se basan en RDF y proporcionan lenguaje para definir ontologías estructuradas, que permiten la integración y la interoperabilidad de los datos entre las comunidades descriptivas.

3.4.1.3. OWL: *Web Ontology Language*

El Web Ontology Language (OWL) como estándar internacional para la codificación e intercambio de ontologías, se basa en lógicas descriptivas que favorecen el desarrollo de aplicaciones semánticas.

Desde un punto de vista técnico, OWL está diseñado para ser utilizado por aplicaciones que necesitan para procesar el contenido de la información en lugar de sólo presentar la información a los seres humanos. Provee axiomas que dotan de restricciones a las clases definidas y las relaciones permitidas entre estos elementos. Como lenguaje para la construcción de ontologías permite realizar especificaciones a niveles abstractos (conceptual) del conocimiento de un dominio. Además utiliza la lógica descriptiva (Ruiz, 2008).

A decir de algunos autores "OWL, el que es sin duda el lenguaje que mayor significado semántico (...) así como el que mayor libertad otorga a la hora de **implementar una ontología**" (Castro Fernández, 2008) y como consecuencia al CMS vinculado a ella.

La capacidad del CMS semántico para vincular información de múltiples fuentes es una característica deseable y potente que se puede utilizar en muchas aplicaciones, dado el poder inferencial de OWL.

3.4.2. S-CMS y ontologías

La investigación en la Web Semántica ha llevado a la estandarización de los lenguajes específicos de ontologías web. Desde los CMS (particularmente los S-CMS, que cuentan con un soporte nativo semántico como RDF, RDF Schema, OWL, etc., o los que son compatibles con módulos desarrollados bajo este fin) se cuenta con la posibilidad de integración a este tipo de tecnología.

Ejemplo de ello se visualiza en (W3C, 2012):

- RDF y OWL que permite a usuarios y/o desarrolladores definir vocabularios ("ontologías") de términos como parte de la Web Semántica.
- RDF proporciona la base para la publicación y la vinculación de los datos. Diversas tecnologías permiten introducir datos en los documentos (RDFa, GRDDL) o exponer lo que tiene en las bases de datos SQL, o ponerlo a disposición en formato RDF.

Otro de los saltos cualitativos en la vinculación de los CMS y las ontologías es la posibilidad de la reutilización de la información con la reutilización de vocabularios existentes para aprovechar la semántica y habilitar agentes de terceros para razonar sobre sus datos. La capa de la ontología permite modificar el modelo sin cambiar la base de datos o volver a escribir ningún código. Esto hace que sea más fácil la edición desde los CMS y se asegura de la estructura de la página web, derivando en una de las principales características de los S-CMS, desde donde se puede gestionar de forma independiente diseño, contenido y semántica.

3.5. Preservación de la información desde los S-CMS

Cuando se abordan los procesos de circulación y preservación de los documentos y la información que existen en la Web, a criterio de Tramullas, puede advertirse que el ciclo de vida de los documentos muestra significativos cambios sobre el ciclo de vida tradicional, al igual que el concepto de documento. **A decir del autor**, "las actividades informativo-documentales deben reformularse para hacer frente **al nuevo medioambiente digital**" (Tramullas, 2005).

Para este nuevo escenario, especialistas como Marcos plantean que, a diferencia de los mecanismos establecidos para el tratamiento tradicionalmente aplicado a documentos en soportes tangibles, "la planificación de la preservación debe comenzar al inicio de la vida de los documentos, que en el caso de los documentos electrónicos se remonta a la concepción del sistema. Si en ese momento se diseña un plan de preservación, durante el resto del ciclo tan sólo será necesario realizar un seguimiento para asegurar que se **cumplen las condiciones establecidas al inicio**"(Marcos Mora, 1999).

La UNESCO reconoce que "la complejidad de los problemas que se plantean obliga a que en la tarea de preservación intervengan los productores de la información digital, comprendidos los de programas informáticos, quienes, al diseñar sus productos, deberán tener en cuenta la conservación. Es obvio que ya han pasado a la historia los días en que la responsabilidad de la preservación incumbía exclusivamente a las instituciones encargadas de los **archivos**" (Llueca, 2003).

La preservación de la información digital ha sido estudiada y conceptualizada por varios especialistas, desde distintas áreas de conocimiento. La mayoría sostiene que el término responde a acciones encaminadas a garantizar "la capacidad de mantener disponibles los ficheros y documentos digitales para que puedan trascender a corto o largo plazo los avances digitales sin riesgo de **que haya alteración o pérdida de legibilidad de los mismos**"(Hernández-Pérez and García-Moreno, 2002).

Termens señala que la preservación digital tiene tres objetivos fundamentales: archivar los documentos digitales, a partir de una adecuada gestión documental; preservar el *bit stream* (es decir, la información en sí misma) y garantizar el acceso a largo plazo (Termens, 2011).

La preservación de los contenidos bajo el uso de los CMS, como en casi todos sus aspectos, tiene más de una perspectiva de estudio. Las más claras y evidentes dimensiones, a criterio de esta autora, son la informacional y la tecnológica. Ambas se complementan y es de vital importancia que se tomen en cuenta para alcanzar integralidad en el proceso y mejores resultados en el mantenimiento y salvaguarda de la información.

Dimensión informacional

Desde la dimensión informacional, el estudio de la preservación de los contenidos, tiene también un alto componente social que le condiciona, pues la importancia de la conservación de la información es totalmente subjetiva y está sujeta a decisiones humanas, dadas a partir de la interacción social, condiciones históricas, económicas y un fuerte componente cognitivo. Se definen a partir de principios de la archivística, la gestión documental o la archivística integrada. Tiene que ver con bases teóricas que advierten formas de proceder para evitar la pérdida de la información y la conservación de sus valores.

La ausencia de normas, pautas y la excesiva dependencia de la tecnología, tanto en el proceso de gestión de preservación de la información digital como en los de creación de productos resultantes de la utilización de un CMS, plantea, desde la dimensión informacional en el análisis del tema, un conjunto de desafíos a considerar.

A criterio de esta autora, al análisis se puede dar en cuatro dimensiones⁶⁰:

⁶⁰Estas conclusiones están basadas en el artículo de: SALVADOR BENÍTEZ, A. & RUÍZ RODRÍGUEZ, R. 2005. Metadatos para la preservación de colecciones digitales. *Cuadernos de documentación multimedia*, 48-60.



Fig. 3.5. Dimensiones de análisis de la preservación de la información digital (elaboración propia)

- Teórico – metodológica: Referida a la importancia de la toma de decisiones en el proceso de preservación de la información digital. ¿Qué elementos se deben tener en cuenta para determinar qué se conserva y qué no? Implica estrategias para garantizar la reutilización de la información en el futuro. Se basa en principios de selección y expurgo tomados del campo de la gestión documental y la archivística. Toma en cuenta las fases del ciclo de vida de la información.

Hernández Quintana, en defensa de criterios interdisciplinarios para abordar el estudio de la información y sus complejidades en el contexto digital, apuesta por la efectividad de las herramientas que aporta esta área de conocimiento para el tratamiento de la información digital. La “unicidad, autenticidad y preservación documental, como principios de la nueva archivística o archivística integrada, enfrentan los retos de la ubicuidad digital, de la manipulación dolosa y la obsolescencia tecnológica” (Bosch, 2001).

- Tecnológica: Determinada por estrategias trazadas en base al nivel de obsolescencia del hardware y el software.

Al respecto, autores como Salvador y Ruiz plantean que “la velocidad con que la tecnología está evolucionando hace casi imposible conservar

programas y sistemas de almacenamiento así como garantizar la reparación, el mantenimiento y la utilización de los equipos. Aun así, esto no sería suficiente si ya no podemos utilizar los programas o peor aún, si no sabemos cuáles se utilizaron” (Salvador Benítez and Ruíz Rodríguez, 2005).

Desde esta dimensión deben considerarse, además, las bondades que ofrece el software para la preservación.

- Socio-económica: Estipula niveles mínimos de compromiso para conservar la información, tanto de la institución como de las personas encargadas de la custodia de la información, mantenimiento mínimos de sus valores (autenticidad⁶¹ e integridad⁶²) y difusión. Para ello es necesario balancear las variables costo y ganancia.

Al respecto Serra plantea que la preservación del material digital es un proceso continuo y a la inversión inicial de creación o de conversión se debe añadir el coste del trasvase de datos cada cinco o diez años como mínimo, tarea que supone no sólo una inversión económica sino también de tiempo y recursos humanos (Serra Serra, 2002).

- Jurídica: En los procesos desarrollados en pos de la preservación digital, debe velarse por el cumplimiento y garantía de los derechos de propiedad intelectual y el derecho a la paternidad, tanto de los productores del material electrónico como del acceso y uso que se hace en el entorno digital por parte de las instituciones y usuarios en general.

Aunque para ganar claridad en la exposición de las ideas, dichas dimensiones se presentan de manera aislada, es válido señalar que su análisis en la práctica

⁶¹Las Directrices de la UNESCO (2003), define la *autenticidad* como “**garantía del carácter genuino y fidedigno de ciertos materiales digitales, es decir, de que son lo que se afirma de ellos, ya sea objeto original o en tanto que copia conforme y fiable de un original, realizada mediante procesos perfectamente documentados**” (BNA 2003. Directrices para la preservación del Patrimonio Biblioteca Nacional de Australia: División de la Sociedad de la Información. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura).

⁶²Las Directrices de la UNESCO (2003), define la *integridad de objetos digitales* como “**estado de los objetos que se encuentran completos y que no han sufrido corrupción o alteración alguna no autorizada ni documentada**” (ibid.).

debe estar matizado por la integración y la interrelación, por el grado de imbricación y dependencia que existe entre ellas.

Otra manera de ver la conservación de la información es a través de la asignación de metadatos de diferentes tipos.

Lazinger y García y Caballero clasifican los metadatos en tres categorías de acuerdo con la naturaleza de los datos que describen (Lazinger, 2001); (García and Caballero, 2009).

- Metadatos descriptivos: describen información bibliográfica. Son aquellos que describen e identifican los recursos de información, facilitan la búsqueda y recuperación. Los que aplican los estándares DublinCore, MARC, etiquetas meta de HTML.
- Metadatos estructurales: referidos a formatos y estructuras. Son aquellos que proporcionan la información sobre la estructura interna de los recursos y la relación entre sus elementos o archivos. Contribuyen a facilitar la navegación y la presentación. Los aplican los estándares SGML, XML/RDF y EAD.
- Metadatos administrativos o de conservación: son aquellos que aportan información relativa a la gestión de derechos, uso del contenido digital, información sobre la autenticidad, historia de la custodia del recurso, estructura y características técnicas.

Anne J. Gilliland establece cinco tipologías (Gilliland, 2008):

- Metadatos administrativos: empleados en la gestión y administración de colecciones y recursos de información. Por ejemplo, información de localización, criterios de selección para la digitalización, acceso legal a los documentos, derechos y reproducción, información de adquisición.
- Metadatos descriptivos: para identificar y describir colecciones y recursos de información relacionados. Por ejemplo, para diferenciar entre versiones de documentos, anotaciones sobre los creadores, relaciones de hiperenlace entre recursos relacionados.

- Metadatos de preservación: relacionados con la gestión de la preservación de las colecciones y recursos de información. Información relacionada con las condiciones físicas del documento, documentación de la preservación de las versiones físicas y digitales de los recursos por ejemplo, fecha de actualización y de migración, las versiones de los documentos, documentación relacionada con los cambios en el documento durante la digitalización y la preservación.
- Metadatos técnicos: relacionados con el comportamiento y las funciones del sistema, por ejemplo, información sobre hardware y software, información técnica de la digitalización, formatos, compresión, autenticación y seguridad de los datos, claves de encriptación, claves y contraseñas.
- Metadatos de uso: relacionados con el nivel y tipo de uso de las colecciones y los recursos, por ejemplo, los registros de circulación, registros de las versiones físicas y digitales, reutilización del contenido.

Como puede advertirse, aunque se den diferentes clasificaciones de metadatos, en la práctica su uso ha conducido la ampliación del alcance de los existentes y/o creación de nuevos metadatos para describir las características de los objetos digitales. Los esquemas de metadatos se han desarrollado, y actualmente existen numerosos esquemas de descripción y tipologías diversas para adaptarse a las características de los recursos que describen (Vaishnav and Sonwane, 2007).

Para el caso de los CMS semánticos, la asignación de metadatos ocurre, al igual que en los CMS genéricos, en diferentes momentos (Fig.3.9.):

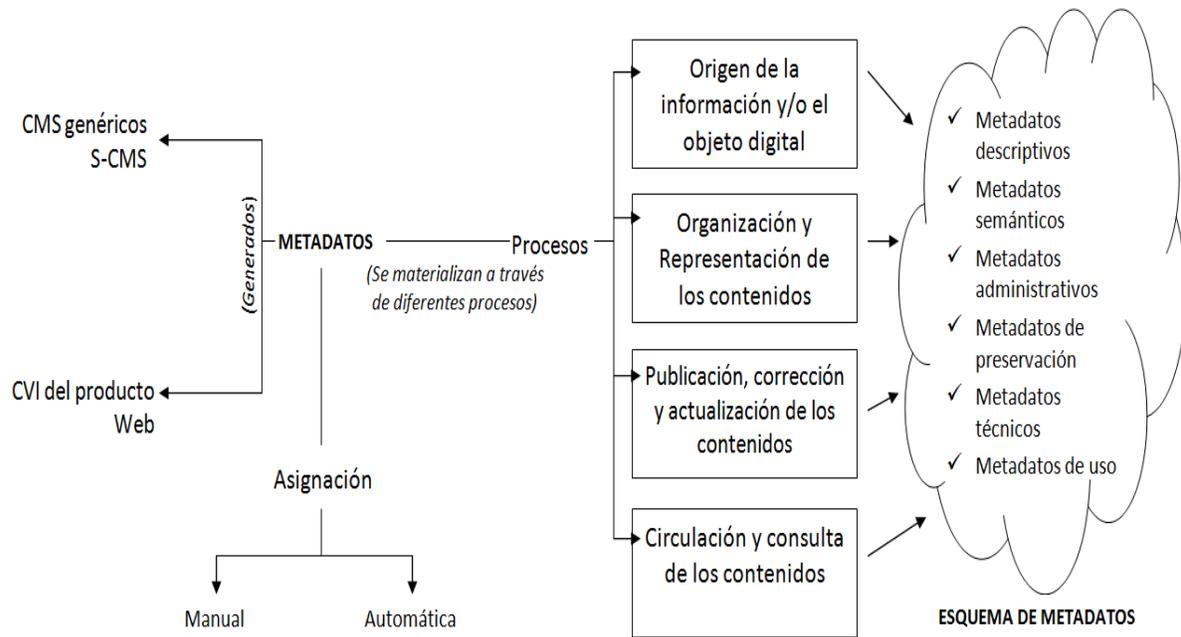


Fig. 3.6. Asignación de metadatos a los contenidos (elaboración propia)

A los contenidos le son asignados un conjunto de metadatos, como se señaló con anterioridad, por dos vías. La primera, de manera automática por el CMS como software; la segunda, metadatos que le son asignados de manera manual o automática una vez que el contenido comienza su ciclo de vida (CVI) en el producto Web. Por tanto, la asignación puede darse en diferentes momentos. En el esquema se representan los cuatro que esta autora identifica como esenciales.

- **Proceso de creación y origen de la información:** se puede dar de manera natural cuando el documento nace en ambiente digital o a partir de procesos de digitalización cuando el documento nace en soportes tangibles. Se asignan metadatos de creación.
- **Procesos de organización y representación de los contenidos:** se asumen los procesos de catalogación, indización, resumen y clasificación definidos por (Vizcaya Alonso, 1997), en su propuesta de modelación de ciclo de vida de la información (CVI). Se asignan metadatos descriptivos (Osuna Alarcón and De la Cruz Gómez, 2010) y dentro de ellos los metadatos semánticos y las conexiones ontológicas de los contenidos.

- *Procesos de publicación, corrección o actualización de los contenidos*: son procesos de creación de enlaces a los contenidos dentro del producto Web asumido como sistema de información. La publicación de contenidos en la más amplia dimensión del término. Se asignan metadatos administrativos y estructurales (Osuna Alarcón and De la Cruz Gómez, 2010).
- *Proceso de circulación y consulta de los contenidos*: referido a la ubicación de los contenidos y documentos dentro de la base de datos del CMS y en el producto Web y los cambios que en este sentido puedan producirse. También implica la frecuencia de uso y consulta que en ocasiones queda registrado en el sitio e implica otras formas de acceso a la información. Se asignan metadatos administrativos (Osuna Alarcón and De la Cruz Gómez, 2010).

Vale advertir que las conexiones ontológicas establecidas entre los metadatos serán constantemente cuestionadas, actualizadas y/o reutilizadas en correspondencia con la evolución léxico, semántica y conceptual del dominio que lo utiliza, como consecuencia lógica del desarrollo dialéctico del conocimiento humano.

La combinación de estos elementos en la asignación de metadatos y el respeto a la sinergia (fig.3.9.), hace que el sistema genere una estructura única para cada contenido que responde a principios archivísticos (unicidad y autenticidad) de máxima importancia en la conservación de la información digital.⁶³

Para el caso particular de los S-CMS, dentro de los metadatos destinados a la descripción, cobran especial importancia los semánticos u ontológicos. Ellos permiten asignar significado a la información, favoreciendo el proceso de integración con vocabularios y con diferentes perspectivas de datos. Con los S-CMS en la construcción de la Web semántica la idea es "añadir metadatos

⁶³Vale aclarar que para la conservación de la información digital se deben tomar otras medidas de carácter práctico, como es la creación de copias de seguridad, protección a partir de la aplicación de antivirus, adecuado mantenimiento y manejo del hardware, entre otras.

semánticos y ontológicos a la información disponible en la Web describiendo información adicional como el contenido, el significado y la relación de los datos. Esta visión deriva del planteamiento inicial de Tim Berners –Lee en el que la Web es un medio universal de intercambio de conocimiento” (Larrinaga et al., 2013).

Existe amplia variedad de metadatos semánticos destinados a describir contenido (por ejemplo: keywords, Classification, description), autoría (por ejemplo: author, owner, publisher), lenguaje (por ejemplo: lenguaje, content-language, traducción), fecha, etc. (Hernández-Pérez et al., 2009).

El esquema de metadatos y su adecuada gestión, cobra especial valor desde los S-CMS por la nomenclaturalización habitual que hace de cada uno de los elementos y por la develación del contexto de las relaciones.

Particularmente bajo la visión de la gestión de contenidos de la web semántica, en la preservación es trascendental la integridad de las conexiones ontológicas que garantizan la interpretación de la información. Ello implica la explicitación o materialización de la elucidación conceptual del pensamiento de un dominio dado. Así, es importante no solo la protección y resguardo de los objetos digitales, también debe velarse por los metadatos y las relaciones establecidas entre ellos, tomando en cuenta el carácter flexible de dichas conexiones y el contexto histórico – profesional desde donde fueron establecidas.

Dimensión tecnológica

La preservación de objetos digitales requiere de una continua supervisión, copias de seguridad, reorganización de ficheros, verificación, etc. (Keefer and Gallard, 2007).

Desde la dimensión tecnológica la preservación de los contenidos viene a dar al traste con el funcionamiento de los CMS como herramienta, las bondades y funcionalidades que permiten que se conserve la información, respetando los principios de autenticidad y unicidad de los contenidos.

En el proceso de circulación de los contenidos en los productos generados sobre la base de un CMS, el mismo objeto puede ser almacenado en varios

formatos digitales o del mismo objeto pueden existir diferentes versiones, con diferencias mínimas o sustanciales. Hay diversidad de permisos, derechos y formas de acceder al contenido. De aquí la importancia de la asignación de metadatos a los contenidos y documentos una vez dentro del sistema, que le permitirán su posterior recuperación.

A ello hay que agregar un conjunto de funcionalidades o bondades del CMS que registran el ciclo de vida de los documentos o contenidos una vez entran al sistema. De esta manera, el software permite realizar un grupo de acciones que tributan a la preservación, siempre y cuando se haga un uso estratégico, consciente y responsable de ellas: Conversión de formatos, control de versiones, bloqueador de registro y funciones de administración.

Estas tareas fueron explicadas por Boldú (Boldú, 2011)

Conversión de formatos

- Se utiliza, como el término sugiere, para convertir documentos de diferentes formatos en componentes o contenidos del gestor mediante reglas básicas. No obstante, los autores deberán revisar que el contenido del documento se haya transformado correctamente, y si es necesario, editarlo y corregirlo.
- Algunos CMS, disponen de la funcionalidad de importar y al mismo tiempo convertir a contenidos del sistema todo un directorio e incluso, toda una web. Funcionalidad muy útil en caso de migraciones.
- Poner funcionalidades concretas

Control de versiones

- Funcionalidad fundamental de un CMS. El sistema guarda las diferentes versiones de un mismo documento, ya sea versiones del flujo de trabajo (borradores, no validadas) o versiones publicadas (la actual y las antiguas).
- Para cada versión se conserva la fecha y el editor para poder identificar quién ha hecho los cambios y cuándo.

- Necesario para el correcto funcionamiento de los flujos de trabajo: hay que conservar las versiones (borradores) de un contenido hasta que sea aprobado y publicado.
- Algunos CMS conservan también versiones de los documentos importados al sistema: PDF, imágenes, etc. (véase fig.10)

Explorador > Història > Comparació de fitxers

Comparació de fitxers

Recurs

Ruta: /universitat/Web/ca/estructura/administrativa/admmar.html
 Títol: Campus del Mar

Attributes

Attributes (3) [Show Identical Attributes](#) [View version 2](#) [View version 4 \(En línia\)](#)

I	Status	Property Name	Version 2	Version 4 (En línia)
1	changed	Mida	5749	5761
2	changed	Última data modificació	09/07/09 10:02	23/10/09 10:07
3	changed	Última modificació per	Admin	u53951

Content

Content (7) [Show Identical Elements](#) [Compare whole file](#)

I	Status	Locale	Element Name	Version 2	Version 4 (En línia)
1	changed	ca	Direccio[1]	Universitat Pompeu Fabra <a ...	<p>Universitat Pompeu Fabra <a ...
2	changed	ca	Funcions[1]	<p>Funcions:</p><p>El Campus del ...	<p>Funcions:</p><p>El camp
3	changed	en	Direccio[1]	Universitat Pompeu Fabra <a ...	<p>Universitat Pompeu Fabra <a ...
4	changed	en	Funcions[1]	<p>Funcions:</p><p>The Mar Campus ...	<p>Funcions:</p><p>The Me
5	changed	en	Nom[1]	Mar Campus	Mar campus
6	changed	es	Direccio[1]	Universitat Pompeu Fabra <a ...	<p>Universitat Pompeu Fabra <a ...
7	changed	es	Funcions[1]	<p>Funciones:</p><p>El campus del ...	<p>Funciones:</p><p>El carr

Fig. 3.7. Ejemplo de control de versiones (Boldú, 2011)

Bloqueador de registros

- Funcionalidad que evita que otros autores, editores o administradores puedan modificar un contenido que está utilizando un autor, editor o administrador concreto.
- El sistema debe dar información sobre el usuario que tiene bloqueado un contenido o un directorio de la estructura de registros.
- Algunos CMS disponen de un dispositivo que avisa a los usuarios cuando un contenido bloqueado deja de estarlo.
- Si el sistema no tiene esta funcionalidad se crearán versiones diferentes o, en el peor de los casos, modificaciones paralelas.

Funciones de administración

- Funcionalidades relacionadas con la capacidad de creación de diferentes tipos de contenido y plantillas de visualización; capacidad de poder controlar la seguridad del sistema, la gestión de usuarios y la asignación de roles, los derechos de acceso y edición a un determinado contenido o estructura, etc.
- Permite la descentralización de la administración de permisos en determinadas secciones a los responsables de esas secciones (imprescindible en contextos grandes y con edición descentralizada).
- La gestión de enlaces comprende diferentes funcionalidades, no siempre todas presentes: librerías de enlaces gestionada por los administradores, herramientas para modificar enlaces a todo el sistema cuando un enlace externo ha cambiado o ha dejado de existir, gestión de enlaces rotos (Normalmente, sólo los internos. El sistema modifica automáticamente los enlaces a contenidos internos cuando éstos cambian de nombre o de lugar en la estructura), funcionalidad que informa a los autores sobre los enlaces entrantes y salientes de un contenido determinado.

En el tratamiento de la información en los CMS se enfoca desde dos caminos: (1) desde las consecuentes implementaciones tecnológicas que los definen y (2) desde el reconocimiento de los principios teóricos que permiten la proyección consciente del trabajo. Esta última, indiscutiblemente, es campo de acción de varias áreas de conocimiento, entre las que se encuentra la CI y sus profesionales.

Dadas las bondades de socialización y democratización de proceso de actualización de la información en los sitios Web contruidos con CMS, los profesionales de la información deben preocuparse por definir, en este nuevo contexto de actuación, sobre quién (o quiénes) recae la responsabilidad de velar por la conservación de la información y objetos electrónicos en general. Cómo deben guardarse los objetos digitales para garantizar la preservación digital a largo plazo.

Sobre este particular la UNESCO (2003) en las *Directrices para la preservación del patrimonio digital*, advierte un conjunto de preguntas que sirven de guía a

los profesionales de la información en la creación de estrategias para la toma de decisiones en el campo de la preservación de materiales digitales:

- **“¿Para quién debe conservarse este material?**
- ¿Por qué vale la pena conservar los materiales?
- ¿Depende el valor de la apariencia del material?
- ¿Depende el valor del modo de funcionamiento del material?
- **¿Depende el valor del contexto del material?”** (Llueca, 2003).

Las respuestas a estas interrogantes, no la encontraremos en las herramientas tecnológicas, sino en las políticas institucionales, creadas con el objetivo de salvaguardar su patrimonio documental.

3.6. Referencias bibliográficas

- Biblioteca Nacional de Australia. 2003. Directrices para la preservación del Patrimonio Biblioteca Nacional de Australia: División de la Sociedad de la Información. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
- BOLDÚ 2011. Curso: "Sistemas de gestió de continguts Digitals (SGCD): Tema 4. TIPUS DE CMS". En: *Màster en Gestió de Continguts Digitals (curso 2010-2011)*. Barcelona, España: [notas de clase].
- BOSCH, M. 2001. Documentos y lenguaje de marcado: conceptos, problemas y tendencias. *El profesional de la información*, 10, 4-9.
- BRATSAS, C., BAMIDIS, P., DIMOU, A., ANTONIOU, I. & IOANNIDIS, L. 2012. Semantic CMS and wikis as platforms for Linked Learning. In: *2nd Int. Workshop on Learning and Education with the Web of Data-24th Int. World Wide Web Conference*.
- CALDÓN, E. F., URIBE, G., LÓPEZ, D. M., DE OLIVEIRA, J. P. M. & WIVES, L. K.n 2010. Mecanismos de anotación semántica de Contenidos en Plataformas de Redes Sociales. *Cadernos de Informática*, 5, 89-99.
- CASTELLS, M. 2002a. La dimensión cultural de Internet. Cultura XXI. [Online] Available: <http://www.uoc.edu/culturaxxi/esp/articles/castells0502/castells0502.html> [Accessed 30 de septiembre de 2013].
- CASTELLS, P. 2002b. Aplicación de técnicas de la web semántica. In: *Workshop de investigación en entornos de interacción colectiva (COLINE'02)*. Granada, España. [Online] Available: http://moodle2.unid.edu.mx/dts_cursos_mdl/pos/TI/BE/AM/12/tecnicas_de_la_web_semantica.pdf [Accessed 23 de abril de 2016]
- CASTRO FERNÁNDEZ, R. 2008. Representación del Conocimiento. Web Semántica. *Universidad Carlos III de Madrid*. [Online] Available: <http://www.it.uc3m.es/jvillena/irc/practicas/08-09/05.pdf> [Accessed 19 de septiembre de 2015]
- CHRIST, F. 2012. Interactive Knowledge Stack A Software Architecture for Semantic Content Management Systems. In: PADERBORN, U. (ed.) *Architekturen 2012*. Paderborn.
- DACONTA, M. C., OBRST, L. J. & SMITH, K. T. 2003. *The Semantic Web: a guide to the future of XML, Web services, and knowledge management*, John Wiley & Sons.
- GARCÍA, N. E. & CABALLERO, S. D. 2009. Metadatos: necesidad e importancia de integrar estándares. *Actas del II Encuentro de Catalogadores*. Biblioteca Nacional de Argentina.

- GILLILAND, A. J. 2008. Setting the Stage. *In*: BACA, M. (ed.) *Introduction to metadata*. 2nd ed.
- HERNÁNDEZ-PÉREZ, T. & GARCÍA-MORENO, M. A. 2002. Intranets y preservación digital, algo más que tecnología. *Revista Andaluz de Patrimonio Histórico*, 237 – 243.
- HERNÁNDEZ-PÉREZ, T., RODRÍGUEZ-MATEOS, D., MARTÍN-GALÁN, B. & GARCÍA-MORENO, M. A. 2009. El uso de metadatos en la administración electrónica española: los retos de la interoperabilidad. *Revista española de documentación científica*, 32, 67-91.
- KEEFER, A. & GALLARD, N. (eds.) 2007. *La preservación de recursos digitales: El reto para las bibliotecas del siglo XXI*: UOC.
- LARRINAGA, F., LIZARRALDE, O., SERNA, A. & GERRIKAGOITIA, J. K. 2013. Caso de uso de open data y linked data en Turismo. *tourGUNE Journal of Tourism and Human Mobility*, 31-39.
- LAZINGER, S. S. 2001. *Digital Preservation and Metadata: History, Theory, Practice*, ERIC.
- LEIVA MEDEROS, A. A. 2011. *Texminer: Un Modelo para el Resumen Automático y la Desambiguación de Textos Científicos en el Dominio de Ingeniería de Puertos y Costas*. Tesis doctoral, Universidad de Granada.
- LIENDL, M., INNOCENTE, V., CASE, M., VAN LINGEN, F., DE ROECK, A., ARCE, P., FÜRTJES, A. & TODOROV, T. 2001. The Role of XML in the CMS Detector Description Database. CERN-CMS-CR-2001-008.
- LLUECA, C. 2003. El proyecto PADICAT (Patrimoni Digital de Catalunya) de la Biblioteca de Catalunya 315 Directrices para la preservación del patrimonio digital [Online] Canberra. Unesco.
- LUNA, J. A. G., PARDO, D. T. & GARCÍA, A. N. L. 2006. *Desarrollo de una ontología en el contexto de la Web semántica a partir de un tesoro documental tradicional*, Universidad de Antioquia.
- MARCOS MORA, M. C. 1999. Los archivos en la era digital. *El Profesional de la Información*, 8(6), 4-13.
- MARTÍNEZ USERO, J. A. 2004. La necesidad de interoperabilidad de la información en los servicios de administración electrónica: xml, una posible solución. [Online] Available: <http://eprints.sim.ucm.es/5653/1/2004-interoperabilidad.pdf> [Accessed 30 de septiembre de 2015]
- MÉNDEZ RODRÍGUEZ, E. M. Year. RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio. *In*: Les biblioteques i els centres de documentació al segle XXI: peça clau de la societat de la informació, 1999. 487-498.

- NAVARRO GALINDO, J. L. 2012. *FLERSA: un sistema semántico de gestión de contenido web (S-CMS)*. Tesis doctoral, Universidad de Granada.
- OSUNA ALARCÓN, M. R. & DE LA CRUZ GÓMEZ, E. 2010. Los sistemas de gestión de contenidos en Información y Documentación. *Revista General de Información y Documentación*, 20, 67-100.
- PASTOR SÁNCHEZ, J. A. & MARTÍNEZ MÉNDEZ, F. J. 2009. Aplicación de tesauros, taxonomías y ontologías en los sistemas de gestión de contenidos mediante tecnologías de la Web Semántica. *Ibersid*, 3,143-153.
- PASTOR SÁNCHEZ, J. A. 2011. *Tecnologías de la web semántica*, Editorial UOC.
- PASTOR SÁNCHEZ, J. A. 2013. Marcado semántico: tecnologías y aplicación para la representación de sistemas de organización del conocimiento en el contexto Linked Open Data. *Scire: representación y organización del conocimiento*, 19 (2), 55-68.
- PASTOR SÁNCHEZ, J. A., ORDUÑA MALEA, E. & SAORÍN, T. 2013. Marcado semántico automático en gestores de contenidos: integración y cuantificación. *El profesional de la información*, 22, 381-391.
- PEDRAZA JIMÉNEZ, R., CODINA BONILLA, L. & ROVIRA, C. 2007. Web semántica y ontologías en el procesamiento de la información documental. *El profesional de la información*, 16, 569-578.
- SALVADOR BENÍTEZ, A. & RUÍZ RODRÍGUEZ, R. 2005. Metadatos para la preservación de colecciones digitales. *Cuadernos de documentación multimedia*, 16(21), 48-60.
- SERRA SERRA, J. 2002. Estrategias de preservación de documentos electrónicos. In: *V Jornadas de Archivos Electrónicos: El reto electrónico: nuevas necesidades, nuevos profesionales*. Priego de Córdoba: National Archives and Records Administration y el Public Record Office. España.
- TERMENS, M. 2011. Conferencia: Buenas prácticas en preservación digital. (Consideraciones inéditas). Barcelona
- TRAMULLAS, J. 2005. Herramientas de software libre para la gestión de contenidos. *Hipertext.net*, 3. [Online] Available: http://www.betsime.disaic.cu/secciones/tec_ja_05.htm [Accessed: 24 de marzo de 2012]
- VAISHNAV, A. & SONWANE, S. S. 2007. Mapping Metadata Standards. *International Conference on Semantic Web and Digital Libraries (ICSD)*. Bangalore.
- VIZCAYA ALONSO, D. (ed.) 1997. *Información: procesamiento de contenido*, Rosario: Nuevo Paradigma.

- W3C. 1999. *Resource Description Framework (RDF): Modelo y Sintaxis Especificación* [Online]. Available: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222> [Accessed 28 de septiembre de 2015].
- WALSH, N. 1998. What is XML. *XML. commune*. [Online] Available: http://www.ce.unipr.it/people/bianchi/Teaching/IntelligenzaArtificiale/rdf_pl/XML-RDF/xmlguide1.html [Accessed 8 de abril de 2016]

CAPÍTULO IV

MARCO METODOLÓGICO DE LA INVESTIGACIÓN

CAPÍTULO IV: Marco metodológico de la investigación

En el presente capítulo se pretende exponer detalladamente los requerimientos y pautas metodológicas que sustentan el estudio para el diseño y propuesta de un CMS semántico para los sitios web de la Universidad de La Habana.

La investigación que se presenta ha sido desarrollada utilizando diversas técnicas y procedimientos. Construir desde la academia un modelo de CMS semántico siempre constituye un reto por las expectativas del dominio para el cual se construye. Este capítulo inicial de la tesis doctoral se centra en la explicitación de aspectos medulares para la comprensión del estudio llevado a cabo en esta investigación.

4.1. Contexto de investigación

La investigación se despliega dentro del marco de la Universidad de La Habana (UH).

La UH fue fundada en 1728. Es la más antigua del país. Actualmente cuenta con 18 facultades, 32 licenciaturas de diferentes ramas del saber, 13 centros de investigación y/o estudio, 30 bibliotecas o centros de documentación y otras estructuras de funcionamiento⁶⁴. Posee varios sistemas oficiales cuya función esencial es la gestión de información.

⁶⁴ La estructura y funcionamiento de la UH y sus sistemas se explicarán con más detenimiento en el capítulo 5, dedicado al abordaje del marco lógico de la investigación

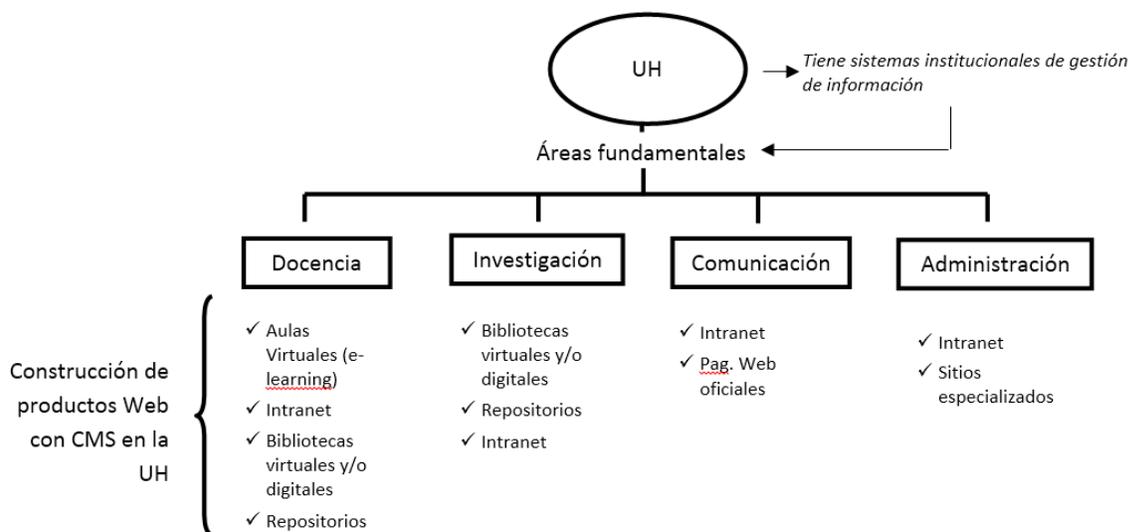


Fig. 4.1. Tipologías de productos Web construidos en la UH (elaboración propia)

La utilización de la Web e internet, son fundamentales para su funcionamiento y dinámica de gestión. No solo con fines investigativos, también implementan y usan las páginas institucionales para difundir informaciones sobre sus capacidades, funciones, lineamientos generales, contactos, etc., es decir, a los elementos que definen su imagen corporativa y funcional.

Por otra parte, la UH surge y se mantiene como universidad pública, característica reforzada dado el sistema político socialista que rige el país, por ende sus procesos esenciales de sostenibilidad son subvencionados por el gobierno. Eso hace que el concepto de racionalización de los recursos disponibles y ahorro sean principios de trabajo bien instaurados en los esquemas laborales de la universidad, lo que unido a las potencialidades reales de los CMS de tipo software libre, ha hecho que en la última década el uso de dichas herramientas se haya popularizado en el contexto universitario cubano y particularmente en la UH.

Las Intranet son implementadas para el funcionamiento interno de cada centro y van estrechamente vinculadas a la gestión estratégica de la información dentro de la institución. Están llamadas a convertirse en el principal y más importante sistema de manejo de la información en línea en las universidades.

Sobre las Intranet, autores como Tramullas, entre otros, llaman la atención sobre los siguientes aspectos:

Acceso, intercambio de información y servicios: en las Intranet, los recursos de información y los servicios disponibles son privados y de uso solamente de los integrantes de la organización, constituyendo plataformas internas de comunicación y espacio virtual para la gestión de información de las organizaciones. El diseño del acceso, intercambio de información y servicios se realiza en función de usuarios previamente identificados.

El usuario debe encontrar lo que necesita conocer con respecto a las funciones que desarrolla la organización, ya que la información se encuentra mucho más organizada y seleccionada, en función de los objetivos, misión y visión. Tramullas **señala que "esto se realiza de forma que resulte completamente transparente para el usuario, pudiendo éste acceder, de forma individual, a todo el conjunto de recursos informativos de la organización, con un mínimo costo, tiempo y esfuerzo"** (Tramullas, 2006).

Preservación, seguridad y publicación de los recursos: La preservación de los recursos va estar dada por políticas de la propia organización. Además, se establecen diversos tipos de permisos a los usuarios para acceder y publicar la información.

Dichos elementos refuerzan la idea de la necesidad del reconocimiento del significado de los términos y la relación entre ellos, no solo para las personas que lo implementos y/o utilizan también, para las máquinas. Así como la normalización del lenguaje y términos utilizados que tribute de manera directa a los procesos de organización, búsqueda y recuperación de la información.

La intranet de la UH se escoge como contexto investigativo, precisamente por ser el espacio de interacción de un dominio donde se expresan diversas disciplinas y profesionales y la riqueza del discurso es notablemente compleja. La posibilidad de montar la intranet con un CMS semántico sería un impacto significativo para la gestión de información en dicho centro.

4.2. Objeto de investigación

Los Sistemas Gestores de Contenido semánticos son el objeto de investigación de la presente tesis. Demanda del abordaje de otros temas estrechamente relacionados a él como lo son la web semántica (como contexto de su aplicación) y las ontologías (producto auxiliar indispensable para su implementación).

4.3. Definición y operacionalización de las variables

Una variable es una propiedad que puede variar (adquirir diversos valores) y cuya variación es susceptible a medirse (Hernández Sampieri et al., 2010). Cada una se contextualiza y explicita, como parte de una investigación, a través de su definición conceptual y su definición operacional.

La definición conceptual, a decir de Baptista, Fernández & Hernández Sampieri, **“define el termino o variable con otros términos”**. Pueden ser definiciones de diccionarios o libros especializados (Kerlinger, 2006) y cuando describen la esencia o las características reales de un objeto o fenómeno se les denomina **“definiciones reales” (Reynolds, 1971. Citado por: (Hernández Sampieri et al., 2010))**.

A los efectos de la presente investigación pueden identificarse desde la hipótesis cuatro variables, las que se han definido de la siguiente forma:

- CMS semántico (S-CMS): es aquel CMS que se apoya en tecnologías semánticas combinadas para aportar descripciones de contenido explícito de los recursos de manera que sean entendibles por un software.
- Organización y Representación de Información (ORI): Partiendo del ciclo de la información propuesto por Vizcaya, los procesos que componen la organización y representación de la información y que por tanto la definen, son: clasificación, indización y resumen (Vizcaya Alonso, 1997).
- Búsqueda y Recuperación de Información (BRI): Proceso que permite recuperar, a través de los diferentes puntos de acceso establecidos en la

fase de ORI del ciclo de vida de la información propuesto por Vizcaya (1997), la información necesaria para identificar, localizar y acceder a los documentos (Vizcaya Alonso, 1997).

- Gestión de Información (GI): Proceso de gestión, orientado al establecimiento de políticas de acceso, procesamiento y consumo de información.

Tabla.4.1. Presentación de la variable para la evaluación de las características semánticas de los CMS (elaboración propia)

Variable	Categoría de análisis ⁶⁵	Indicadores	Definición
S-CMS	Características semánticas	Soporte nativo para la gestión de vocabularios y ontologías	Capacidad de gestión de RDF y RDFa como formatos nativos para el manejo de vocabularios como bibo, vivo, geontology, skos, foaf, lom, dublin, core, sioc.
		Asociación entre los contenidos estructurados del portal y los vocabularios RDF gestionados	Posibilidad de asociar cada tipo estructurado a una o varias clases RDF y mapear parte de los campos del tipo estructurado a propiedades RDF y RDFa .
		Adquisición de módulos	Posibilidades de extensión e integración a través de la vinculación a módulos extras extendiendo la funcionalidad básica del CMS con posibilidades de acciones semánticas
		Asociaciones RDF	Posibilidad de publicar las asociaciones RDF creadas en formatos semánticos
		SparQL Endpoints y Linked Data	Posibilidad de manejar datos abiertos o datos enlazados y de realizar consultas a través de SPARQL Endpoint

⁶⁵ El término **categoría de análisis**, para el presente estudio, puede utilizarse como sinónimo de parámetro. Desde el punto de vista del alcance temático es un término que se encuentra entre la variable y los indicadores establecidos para su evaluación.

Tabla.4.2. Presentación de la variable ORI (elaboración propia)

Variable	Categoría de análisis	Indicadores	Definición
ORI	Calidad de los datos	Exactitud	Grado en el cual los datos son certificados, libre de errores, correctos y confiables
		Compleitud	Grado en el cual los datos no se encuentran ausentes o perdidos, cubren las necesidades de las tareas, y son lo suficientemente amplios y profundos para una tarea determinada. No existen valores ausentes para un atributo dado en el sistema
		Consistencia	Grado en el cual la forma y el contenido en los campos de datos es la misma a través de múltiples sistemas fuentes
		Objetividad	Grado en el cual los datos son objetivos, sin prejuicios, basados en hechos e imparciales
		Interpretabilidad	Grado en el cual los datos están en un lenguaje apropiado y las definiciones son claras
		Facilidad de comprensión	Grado en el cual los datos son comprendidos fácilmente
		Consistencia representacional	Grado en el cual los datos son continuamente presentados en el mismo formato
		Representación concisa	Grado en el cual los datos son representados de forma compacta, bien organizados y estéticamente agradables

Tabla.4.3. Presentación de la variable GI (elaboración propia)

Variable	Categoría de análisis	Indicadores	Definición
GI	Calidad de los datos	Tiempo de vida	Grado en el cual la edad de los datos es apropiada para una tarea determinada. Tiempo de caducidad del dato. El tiempo de vida tiene dos componentes: la actualidad y la volatilidad.
		Actualidad	Medida de cuán antiguos son los datos, basada en cuánto tiempo

			hace desde que estos fueron grabados.
		Volatilidad	Medida de cuan inestables pueden ser los datos, o sea, la frecuencia del cambio de un valor para un atributo de entidad
		Veracidad	Grado en el cual los datos son inconsistentes, o sea, se refiere directamente a problemas de calidad de datos. Hace referencia al nivel de fiabilidad asociado a ciertos tipos de datos.
		Credibilidad	Grado en el cual los datos son considerados como verdaderos y creíbles
		Cantidad apropiada de datos	Grado en el cual el volumen de información es adecuado para una tarea determinada
		Accesibilidad	Grado en el cual los datos están disponibles o son fáciles y rápidos de recuperar
		Unicidad	Grado en el cual los valores de datos están restringidos en un conjunto con entradas distintas. Cada valor es único en su tipo
		Seguridad	Grado en el cual el acceso a los datos está restringido apropiadamente para mantener su seguridad
		Utilidad	Cada elemento de los datos en el sistema debe satisfacer ciertos requerimientos de las colecciones de los usuarios
		Preservación	Propiedad de calidad que poseen los datos digitales. Refleja un estado del dato, el cual puede variar debido a cambios de hardware, software, formatos y procesos a los que el dato es sometido (migración)
		Comodidad	Grado en el cual los caminos de navegación resultan difíciles. Número de caminos de navegación perdidos/interrumpidos

Tabla.4.4. Presentación de la variable BRI (elaboración propia)

Variable	Categoría de análisis	Indicadores	Definición
BRI	Calidad de los datos	Credibilidad	Grado en el cual los datos son considerados como verdaderos y creíbles
		Precisión	Número de lugares decimales o dígitos significativos en las medidas. La falta de precisión puede verse como inexactitud.
		Objetividad	Grado en el cual los datos son objetivos, sin prejuicios, basados en hechos e imparciales
		Reputación	Grado en el cual los datos son altamente considerados en términos de sus fuentes o contenidos
		Valor agregado	Grado en el cual los datos son beneficiosos y proporcionan ventajas de su uso. Los datos adicionan valores a las operaciones
		Relevancia	Grado en el cual los datos son aplicables y útiles para una tarea determinada
		Cantidad apropiada de datos	Grado en el cual el volumen de información es adecuado para una tarea determinada
		Interpretabilidad	Grado en el cual los datos están en un lenguaje apropiado y las definiciones son claras
		Facilidad de comprensión	Grado en el cual los datos son comprendidos fácilmente
		Accesibilidad	Grado en el cual los datos están disponibles o son fáciles y rápidos de recuperar
Comodidad	Grado en el cual los caminos de navegación resultan difíciles. Número de caminos de navegación perdidos/interrumpidos		

4.4. Campo de investigación

Urías (2009) señala que el campo de la investigación es el segmento particular más amplio donde cabe el objeto de la investigación, es el campo quien contiene al objeto y suele ser más amplio que este. Partiendo de ello se puede declarar que el escenario investigativo global son los Sistemas Gestores de Contenido vistos como herramientas tecnológicas auxiliares de la Ciencia de la Información.

Sin embargo, como investigación entronca con otros saberes como las Ciencias de Computación, Políticas de la Información, Educación, Derecho. Por tales razones es fácil advertir su matiz multidisciplinar, rasgo distintivo de nuestro campo profesional.

4.5. Tipo de investigación

Autores como Dankhe (1986) y Baptista, Fernández & Hernández Sampieri (Hernández Sampieri et al., 2010), clasifican las investigaciones desde las ciencias sociales en cuatro tipologías atendiendo a su alcance: exploratoria, descriptiva, correlacional y explicativa. En función del tipo de estudio varía la estrategia a seguir en su desarrollo, aunque en la práctica suelen combinarse.

Atendiendo a los criterios de dichos autores, el presente trabajo responde a las características de un estudio descriptivo, pues entre los propósitos del estudio se encuentra proponer y describir un CMS semántico personalizado para los sitios web de la UH. Para ello se hizo necesario durante el periodo investigativo y en el informe escrito dedicar tiempos a análisis teóricos bajo una mirada crítica.

4.6. Enfoque de investigación

Bajo la visión metodológica de Baptista, Fernández & Hernández Sampieri, existen tres posibles enfoques para desarrollar una investigación: cuantitativa, cualitativa o mixta (Hernández Sampieri et al., 2010). El estudio se defiende bajo un enfoque cualitativo esencialmente, aunque se auxilia de elementos

matemáticos y de modelación para la elaboración y presentación de la propuesta de CMS semántico.

Bajo el enfoque cualitativo se “utiliza la recolección de datos sin medición numérica para descubrir o afinar preguntas de investigación en el proceso de **investigación**” (Hernández Sampieri et al., 2010). La investigación se presenta con métodos típicos de este enfoque como: entrevista no estructurada, encuestas y estudio exploratorio de las necesidades enfocadas a la búsqueda y recuperación de información.

4.7. Métodos de investigación

Metodología se refiere al conjunto de métodos y técnicas que se aplican sistemáticamente durante un proceso de investigación para alcanzar un resultado teóricamente válido. En este sentido, la metodología funciona como el soporte conceptual que rige la manera en que aplicamos los procedimientos en una investigación.

4.7.1. Metodologías

Las metodologías para el desarrollo de la investigación se presentan a continuación:

- **“Metodología de la investigación”** de Hernández Sampieri, Fernández Collado y Baptista Lucio para definir y enmarcar las características formales y estructurales de la investigación (Hernández Sampieri et al., 2010).

Es una metodología muy utilizada en el ámbito investigativo por los niveles de alcance generalizadores. Es clara, con un alto número de categorizaciones y descripciones de cada una, por lo que facilita el entendimiento y clasificación de la investigación.

- Para el desarrollo de la ontología se ha hecho uso de la metodología **expuesta en “Desarrollo de Ontologías-101: guía para crear tu primera ontología”** (Noy and McGuinness, 2005) por la simplicidad, claridad y objetividad de los pasos establecidos.

4.7.2. Métodos

Los métodos utilizados en la investigación, a decir de autores (Urías, 2009), (Hernández Sampieri et al., 2010), deben estar en correspondencia con el objeto de estudio y el campo de investigación, por ello se apela a métodos tanto de nivel empírico como teórico.

Teóricos

Los métodos de información de orden teórico son aquellos que se desarrollan a partir del conocimiento de la percepción del ser humano, y se basan en acciones lógicas del pensamiento (Urías, 2009). En la investigación están presente los siguientes:

- ***Lógico-histórico***: Se emplea para estudiar el desarrollo lógico e histórico de los principales criterios sobre el tema de los CMS como herramienta tecnológica de manera general y los CMS semánticos de manera particular. Se Estudian además temas relacionados fundamentales para el desarrollo de la investigación como web semántica, datos enlazados, ontologías, etc. Permitted la construcción de un marco teórico para conocer las particularidades e interioridades del mismo. Se parte de la aparición de los CMS en la segunda mitad del siglo XX. Es utilizado para secuenciar el desarrollo del tema en todos sus aspectos teóricos-conceptuales.
- ***Analítico-sintético***: Posibilita analizar por partes los principales documentos y se utilizó para penetrar en la esencia del fenómeno objeto de estudio y posteriormente realizar una síntesis de la información obtenida. Permite arribar a conclusiones sobre las metodologías estudiadas, así como el desarrollo y uso de la herramienta.
- ***Inductivo-deductivo***: Permite realizar análisis e inferencias de los datos obtenidos permitiendo arribar a las regularidades y conclusiones presentadas a lo largo de la investigación.

Empíricos

- ***Análisis documental clásico***: Este método se basa en la recopilación de información a partir de la consulta de fuentes de información diversas.

Esencialmente es empleado para poder sustentar teóricamente la investigación. Así, través de los métodos teóricos antes explicados, se realizó un minucioso análisis documental, sin perder de vista el análisis del contexto histórico, lógico y social.

Para el análisis de las bases teóricas que sustentan el uso y aplicación de los CMS a los distintos proyectos, se distinguieron los siguientes pasos:

- Determinar los objetivos de estudio documental: Dirigido a revelar los documentos que existen sobre el tema para realizar, mediante ellos, un trabajo de valoración crítica que refrende este tipo de análisis.
- Establecer la muestra de los documentos que serán estudiados: Se establece una muestra de los documentos que serán analizados para determinar los elementos aspectos fundamentales de los CMS. Podrá advertirse que la variedad de las fuentes, el tratamiento de los temas, las percepciones y las opiniones de los autores referidos tienen una base multidisciplinar y multilingüística, acorde a la pluralidad de elementos de base involucrados. Gran parte de las fuentes estudiadas se encuentran en soporte electrónico o fueron consultadas en Internet.
- Determinar las unidades de análisis en las que se fracciona el contenido para estudiar el documento: Teniendo en cuenta el contexto histórico se analizan paso a paso los períodos de producción editorial representados en los documentos. Se analiza el origen de las primeras aportaciones sobre la temática y el valor de las mismas para la investigación.
- Realizar el estudio documental registrando la información: Se realizó el registro de la información siguiendo la lógica de trabajo expresada en los pasos anteriores, sin perder el análisis contextual, tecnológico y filosófico de esta temática, logrando una integración coherente de todas las ideas para, de forma armónica, entrelazar y sistematizar todos los referentes teórico metodológicos que han rodeado el desarrollo de los CMS.

- Valoración de la información obtenida: Se realizaron valoraciones a partir de los presupuestos teóricos y metodológicos precisados, realizándose inferencias, argumentaciones, críticas y, finalmente, arribando a conclusiones sobre los elementos que debe contener un resumen científico para ser expresado en una nueva metodología.
- **Observación:** para el levantamiento de información sobre los CMS utilizados en la UH para la construcción de sitios web institucionales y sistemas de gestión de información administrativa. El objetivo era pesquisar.

En la presente investigación se desarrolló específicamente una modelación gráfica de lo que va a proponerse como CMS y la relación de sus estructuras. Además se hará una modelación descriptiva de la ontología de la UH, además sus propiedades y relaciones fundamentales de los diferentes procesos.

4.7.3. Instrumentos

- **Entrevista:** Se aplicará a los especialistas que se desempeñen en el uso e implementación de los productos digitales de la UH como parte de la identificación de necesidades y la búsqueda de la factibilidad de la propuesta de un CMS semántico para la UH
- **Encuesta:** Se diseñaron diferentes encuestas, en función de las necesidades investigativas de la autora y el perfil de los usuarios a los cuales se aplicaría la misma:
 - **Encuesta 1:** para especialistas en CMS con el fin de recopilar criterios sobre las características semánticas de los CMS a evaluar. (Ver Anexo 1)
 - **Encuesta 2:** a directivos de la UH con el objetivo de conocer el nivel de conocimiento y conformidad de los usuarios que utilizan los sistemas de gestión en la UH, sobre las posibilidades del software que se emplean para ello. (Ver Anexo 2)

4.7.4. Herramientas

- *Protégé 4.3* para la modelación e implementación de la ontología UH.

Protégé es un software editor de ontologías gratuito y de código abierto creado por la Universidad de Stanford. La instalación es muy sencilla y la configuración por defecto funciona bien. Para trabajar con la versión utilizada se requiere tener un Java, o descargar un instalador que la incluya. Es herramienta útil para crear ontologías en un formato estandarizado como OWL o RDF, permite compartir y utilizar otras aplicaciones y plataformas.

Para el estudio del comportamiento de los recursos utilizados en la tesis se dispuso de los siguientes softwares:

- *EndNote X4*, con vistas a realizar la normalización necesaria de los datos obtenidos y manejo de la bibliografía.
- *MicroSoft Excel*, para la creación de tablas y gráficos correspondientes a algunos de los indicadores.
- *Bibexcel*, para la obtención de matrices en el análisis de los indicadores co-ocurrencia de descriptores y categorías temáticas.
- *Ucinet 6.0* y *NetDraw 1.2*, para la visualización de las matrices elaboradas a través de mapas de redes que facilitan su análisis e interpretación.

Es importante reseñar que las referencias a los recursos citados adoptaron el estilo Harvard

4.7.5. Selección de la muestra

Para el estudio, en función de los objetivos y estrategias de investigación propuestas se han seleccionado diversas muestras, a las cuales se les aplicó diferentes instrumentos.

- Especialistas en el trabajo con los CMS, de manera particular con DRUPAL y JOOMLA, con el objetivo de evaluar las características semánticas de los mismos. En la selección de dichos especialistas, para

evitar el sesgo y el vicio en el proceso evaluativo, se escogen a aquellos que laboran fuera de la UH. Por tanto se convoca a :

- Dr. Félix Oscar Fernández Peña
 - Dr. Alfredo Simón Cuevas
 - MsC. Yuniel Hidalgo
 - MsC. Leandro Tabares
- Usuarios decisores (56) son aquellos que por sus responsabilidades administrativas forman parte del Consejo Universitario de la UH. Eso incluye:
- Equipo Rectoral (10)
 - Directores de áreas centrales (12)
 - Decanos (18)
 - Directores de Centros (12)
 - Otros puestos clave como la Dirección de Informatización (2) y la Dirección de Información (2)
- Dentro del grupo anterior se encuentran los especialistas decisores de implementación de sistemas en la UH, desde la responsabilidad de Director de Informatización y Director de Información, cuya visión integral de la implementación de los CMS en la UH es de vital importancia para la presente investigación.

Por tanto, la muestra seleccionada asciende a 61 personas, distribuidos de la siguiente forma: (Ver tabla 4.5)

Tabla.4.5. Resumen sobre la selección de la muestra (elaboración propia)

Técnica	Muestra de estudio	Muestra
Guía de evaluación (CMS)	Especialistas en CMS (Drupal y/o Joomla)	5 (muestra intencional)
Encuesta (calidad de datos)	Usuarios decisores	56

4.8. Etapas de investigación

La investigación cuenta con las siguientes fases:

- Fase 1: Revisión de la literatura

Revisión de materiales bibliográficos disponibles e identificación referentes teóricos, prestando especial atención a los CMS, S-CMS y la web semántica, así como sus herramientas de construcción y análisis.

- Fase 2: Diagnóstico

Diagnóstico y evaluación de los CMS implementados en la UH

- Fase 3: Diseño del modelo

Construcción del modelo propuesto con la arquitectura, requisitos, componentes estructurales y sistema de recuperación de la información asociado a él.

4.9. Análisis de los recursos de información utilizados en la tesis

En el presente epígrafe se pretende hacer una valoración sobre los recursos de información utilizados en la investigación.

4.9.1. Los indicadores

Desde la Segunda Guerra Mundial se le ha prestado especial atención a la recolección de información y estadísticas con relación a las instituciones y los recursos con vistas a la evaluación y valoración de la actividad investigativa desde la perspectiva científico – tecnológica. Como campo de conocimiento que

se preocupa y ocupa de estos menesteres surge la Bibliometría, definida como “la aplicación de las matemáticas y los métodos estadísticos para analizar el curso de la comunicación escrita y el curso de una disciplina” (Arencibia Jorge, 2010).

En la literatura especializada aparecen definidas distintas tipologías de indicadores. Para el presente trabajo fueron utilizados indicadores unidimensionales y multidimensionales.

Entre los indicadores unidimensionales se encuentran:

Productividad por año: Se mide a partir del conteo de artículos agrupados por el año para determinar cómo se ha comportado el tema. Como resultado, saldrán a la luz los años de mayor y menor productividad, dando los niveles de actualización de la investigación.

Productividad por categorías temáticas: Se mide a partir de la agrupación de los artículos publicados bajo el tema por la categoría temática. La categoría temática es asignada por los administradores y especialistas para hacer una primera clasificación temática de los registros. Esta clasificación es independiente a los criterios de los autores de los artículos. Como resultado se podrán realizar valoraciones sobre los temas trabajados a diferentes niveles en la tesis.

Productividad de autores: Se mide a partir de la agrupación de los autores utilizados en las referencias bibliográficas y en la bibliografía consultada. Como resultado se podrán presentar valoraciones sobre los autores más utilizados y citados en el cuerpo de la tesis.

Países: Se mide a partir de la agrupación de los recursos publicados por países. Este indicador permite observar tendencias geográficas en las publicaciones utilizadas. La supremacía de una región puede sugerir el desarrollo del tema en determinadas zonas geográficas o la necesidad de publicar en determinado idioma, condicionado por las conductas asumidas ante determinadas relaciones de poder que rigen el marco del desarrollo socio-tecnológico y científico.

Tipo documental: Se mide a partir de la agrupación de los trabajos utilizados según su tipología documental.

Los indicadores multidimensionales seleccionados para el estudio son:

Co-ocurrencia de categorías temáticas: Revela las relaciones y frecuencias con que las categorías temáticas asignadas coinciden a la vez en un documento. Puede revelar el nivel de interdisciplinariedad del campo en caso de que se establezcan nexos con otras disciplinas.

4.9.2. Resultado del Análisis métrico

Para la realización de la investigación, por los niveles de complejidad que la envuelven, fue necesario acudir al estudio de varias temáticas, esencialmente a: CMS, CMS semántico y Web semántica, temas cuya imbricación los convierten en el eje teórico que sustenta la tesis. Por dicha razón le fueron dedicados capítulos específicos, antes de la presentación de la propuesta del CMS UH-Web.

El comportamiento de las publicaciones sobre estos temas, han condicionado el uso de la bibliografía consultada del presente estudio. La investigación se ha distinguido por un amplio espectro de temas vinculados a los CMS semánticos. Como resultado se obtuvo un fichero EndNote con 308 registros bibliográficos consultados, de los cuales 251 responden a referencias bibliográficas en el cuerpo de la tesis. Dentro del **fichero "TXT" en el que se registraron y guardaron los resultados**, cada uno de los datos de las publicaciones fue categorizado mediante TAGs, lo que facilitó el análisis a partir de los indicadores previstos.

A continuación se presentan los resultados del estudio realizado en torno al comportamiento de la producción científica desde los recursos de información utilizados en la tesis.

1.9.2.7. Tipo documental

La tipología documental consultada ha sido diversa, lo que indica variedad de fuentes utilizadas en el estudio, en consonancia con las exigencias investigativas para el desarrollo del tema (ver tabla 4.6).

Tabla 4.6. Tipologías documentales utilizadas (elaboración propia)

Tipos documentales	Cant. Registros
Journal Article	174
Web Page	29
Book	22
Generic	19
Conference Proceedings	18
Thesis	11
Report	8
Conference Paper	7
Book Section	6
Electronic Article	3
Electronic Book Section	3
Blog	2
Edited Book	2
Aggregated Database	1
Legal Rule or Regulation	1
Personal Communication	1
Unpublished Work	1
TOTAL	308

Hay un desbalance notable, pero justificado, en los tipos de recursos consultados. Más del 50 % de la bibliografía son artículos de revistas (ver fig. 4.2). Este resultado es reflejo del comportamiento editorial en el ámbito de las tecnologías. El tema, es dinámico, evoluciona rápidamente. Se ha vuelto una práctica habitual que se publique más frecuentemente en revistas y espacios web, que en libros. Como consecuencia, los aspectos más actuales se encuentran en dicho tipo documental.

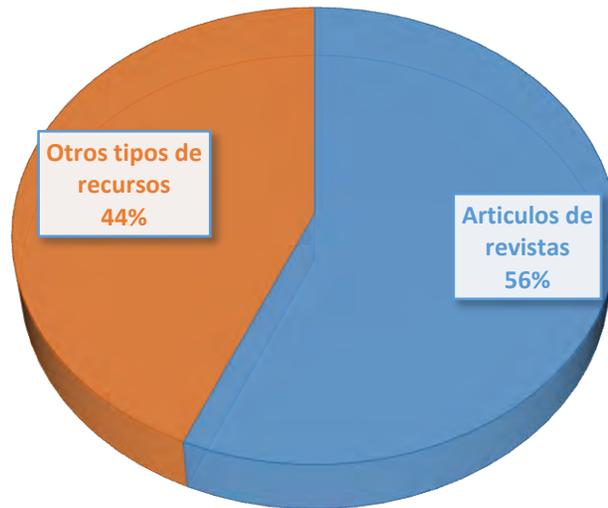


Fig. 4.2. Balance sobre los tipos documentales (elaboración propia)

Además de los artículos, se distinguen otras tipologías con un uso significativo en la investigación: actas de congresos, libros, páginas web, tesis, reportes, ponencias de eventos y secciones de libros (electrónicos en su mayoría) (ver fig 4.3).

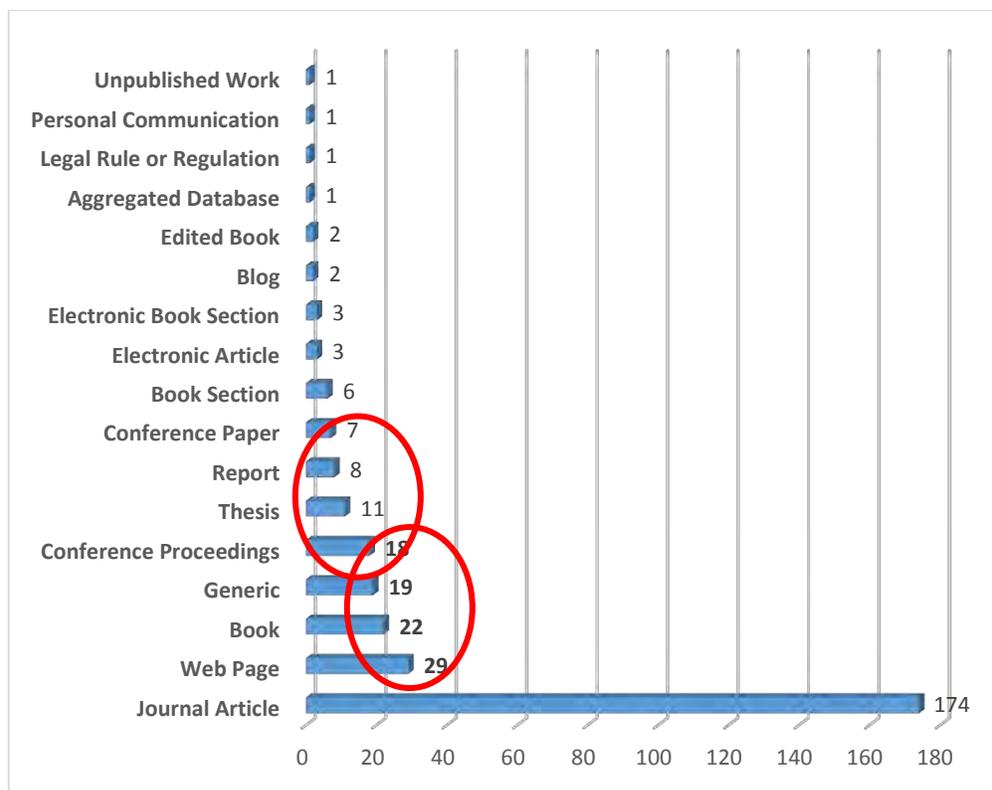


Fig. 4.3. Tipos documentales utilizados (elaboración propia)

1.9.2.1. Productividad por año de publicación

Las publicaciones sobre CMS en el WoS comienzan a aparecer en 1998, fecha próxima a la aparición de este tipo de productos en el mercado. Sin embargo, son escasas las referidas a CMS semántico, donde solo se recuperaron 3 registros. Eso puede estar condicionado a la novedad del tema.

En la tesis los documentos presentan diversas fechas. A continuación se muestran los años de los 308 registros consultados (en uno no se declara el dato) (ver tabla 4.7):

Tabla 4.7. Cantidad de registros consultados por año (elaboración propia)

Año	Cant. Registros	Año	Cant. Registros
1966	1	2003	13
1983	1	2004	19
1987	1	2005	8

1991	1	2006	13
1993	2	2007	14
1994	1	2008	14
1995	3	2009	26
1996	2	2010	17
1997	4	2011	22
1998	6	2012	24
1999	8	2013	25
2000	4	2014	24
2001	9	2015	31
2002	8	2016	6
Total: 307 registros			

Puede advertirse que más del 70% de la bibliografía utilizada responde a los últimos 10 años, lo que habla de un alto nivel en la actualización de los contenidos sobre los que se trabajó la propuesta de CMS semántico (ver fig. 4.4).

Cant. Registros consultados por períodos

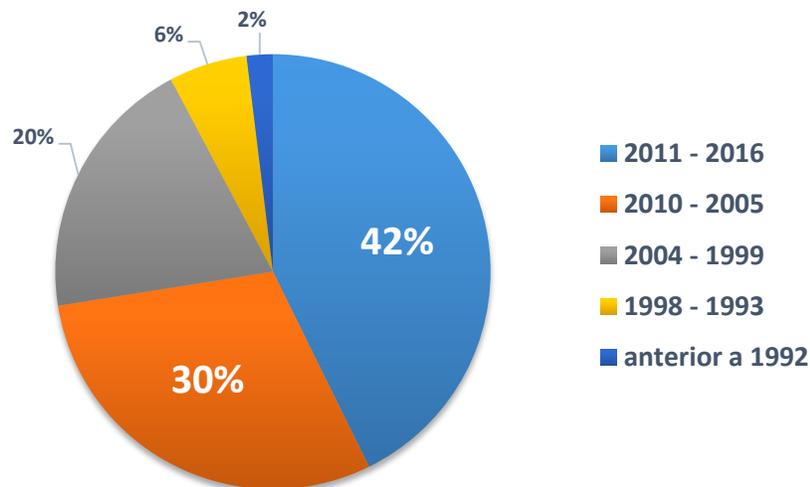


Fig. 4.4. Balance sobre los niveles de actualización de la bibliografía consultada (elaboración propia)

Los registros anteriores a 1992, responden a la literatura clásica de autores claves en el manejo del tema como Gruber y Beners-Lee.

1.9.2.3. Productividad temática y co-ocurrencia a partir de las palabras clave

En la tesis, los documentos mayormente utilizados son los referidos a Web semántica, ontología, datos enlazados y otros aspectos que giran en torno a estos elementos (ver tabla 4.8)

Tabla 4.8. Cantidad de registros consultados por palabras clave (elaboración propia)

Palabras clave	Cant. Registros
semantic web	35
Ontology	14
linked data	8
description logics	7
RDF	7
OWL	6
systems	6
web service	6
Semantic	6
design	5
framework	4
management	4
inference	3
SPARQL	3
Web	3
architecture	2
classification	2
content management	2
Data	2
information	2
Interoperability	2
libraries	2
Linked open data	2
metadata	2
model	2
Reasoning	2
retrieval	2
Semantic Web Services	2
Semantic web technology	2
SWRL	2

Los documentos que explícitamente asumen a los CMS (en cualquiera de sus formas de expresión) son escasos. Esto es consecuencia a que son pocas las publicaciones que sobre CMS con características o bondades semánticas aparecen en la literatura. Esta idea ha sido corroborada con anterioridad con las métricas obtenidas en el WoS.

Por tanto, es fácil deducir que los aspectos semánticos en CMS aportados se obtuvieron, no directamente de citas en el tema sino, de extrapolaciones teóricas tomadas de los aspectos medulares que permiten la construcción de la Web semántica.

1.9.2.5. Productividad de autores y autores más citados

En la bibliografía se identificaron 653 autores, de los cuales 439 fueron citados para apoyar los contenidos y teorías expuestas en la investigación.

Los autores más utilizados en la investigación se muestran en la tabla 4.9:

Tabla 4.9. Autores con mayor cantidad de registros consultados (elaboración propia)

Autores	Cant. Doc.	Autores	Cant. Doc.
Gómez-Pérez, A.	10	Chute, C. G.	2
Patel-Schneider, P. F.	9	Codina Bonilla, L.	2
Bizer, C.	8	Vila-Suero, D.	2
Pastor Sánchez, J. A.	8	Mitchell, S.	2
Horrocks, I.	8	Isele, R.	2
W3C	7	Cyganiak, R.	2
Van Harmelen, F.	6	Decker, S.	2
Parsia, B.	5	Janowicz, K.	2
Tramullas, J.	5	Fensel, D.	2
Motik, B.	5	Kellogg, G.	2
McGuinness, D. L.	5	Fernández-López, M.	2
Corson-Rikert, J.	4	Fokoue, A.	2
Berners-Lee, T.	4	García-Moreno, M. A.	2
Isaac, A.	4	Klyne, G.	2
Lozano-Tello, A.	3	Guzmán Luna, J. A.	2
Rovira, C.	3	Longley, D.	2
Haslhofer, B.	3	Lanthaler, M.	2
Carroll, J. J.	3	Hayes, P.	2
Bechhofer, S.	3	Heath, T.	2
Beckett, D.	3	Heflin, J.	2

Brickley, D.	3	Hendler, J.	2
Corcho, O.	3	Paepcke, A.	2
McBride, B.	3	Ma, Z. M.	2
Cuenca Grau, B.	3	Schandl, B.	2
Ding, Y.	3	Palacios-Madrid, V.	2
Lowe, B. J.	3	Navarro Galindo, J. L.	2
Fernández, M.	3	Schneider, M.	2
Noy, N. F.	3	Sporny, M.	2
Leiva Mederos, A. A.	3	Rosell León, Y.	2
Schreiber, G.	3	Ruttenberg, A.	2
Gruber, T.	3	Senso Ruiz, J. A.	2
Antoniou, G.	2	Studer, R.	2
Bao, J.	2	Matinfar, F.	2
Hernández-Pérez, T.	2	Patil, R.	2
Hidalgo-Delgado, Y.	2	Van de Walle, R.	2
Volz, J.	2	Mosses, P. D.	2
Bidoit, M.	2	Sánchez-Cuadrado, S.	2
Lehmann, J.	2	Zeng, M.L.	2
Bobillo, F.	2	Pedraza Jiménez, R.	2
Miller, L.	2	Tao, C.	2
Carothers, G.	2	Mannens, E.	2
Lausen, G.	2		

Los autores con 2 o más documentos citados se muestran en la tabla 4.10:

Tabla 4.10. Autores con mayor cantidad de registros citados (elaboración propia)

Autores	Cant. Doc.	Autores	Cant. Doc.
Gómez-Pérez, A.	10	Ding, Y.	2
Patel-Schneider, P. F.	9	Fensel, D.	2
Horrocks, I.	8	Fokoue, A.	2
Pastor Sánchez, J. A.	8	García-Moreno, M.A.	2
Bizer, C.	7	Hayes, P. J.	2
W3C	7	Heflin, J.	2
Fernández-López, M.	5	Hendler, J.	2
McGuinness, D. L.	5	Hernández-Pérez, T.	2
Motik, B.	5	Hidalgo Delgado, Y.	2
Parsia, B.	5	Kellogg, G.	2
Isaac, A.	4	Klyne, G.	2
Tramullas, J.	4	Lanthaler, M.	2
Van Harmelen, F.	4	Longley, D.	2
Bechhofer, S.	3	Martínez Méndez, F. J.	2
Beckett, D.	3	Miller, L.	2

Berners-Lee, T.	3	Navarro Galindo, J.L.	2
Brickley, D.	3	Noy, N. F.	2
Carroll, J.	3	Paepcke, A.	2
Grau, B. C.	3	Patil, R.	2
Gruber, T.	3	Pedraza Jiménez, R.	2
Haslhofer, B.	3	Rosell León, Y.	2
Leiva Mederos, A. A.	3	Ruttenberg, A.	2
Lozano-Tello, A.	3	Schandl, B.	2
McBride, B.	3	Schneider, M.	2
Rovira, C.	3	Senso Ruiz, J.A.	2
Schreiber, G.	3	Sporny, M.	2
Bao, J.	2	Studer, R.	2
Bidoit, M.	2	Vila-Suero, D.	2
Codina Bonilla, L.	2	Volz, J.	2
Corcho, O.	2	Zeng, M.L.	2
Cyganiak, R.	2		

En las tablas anteriores se distinguen diferentes autores (ver Fig. 4.5):

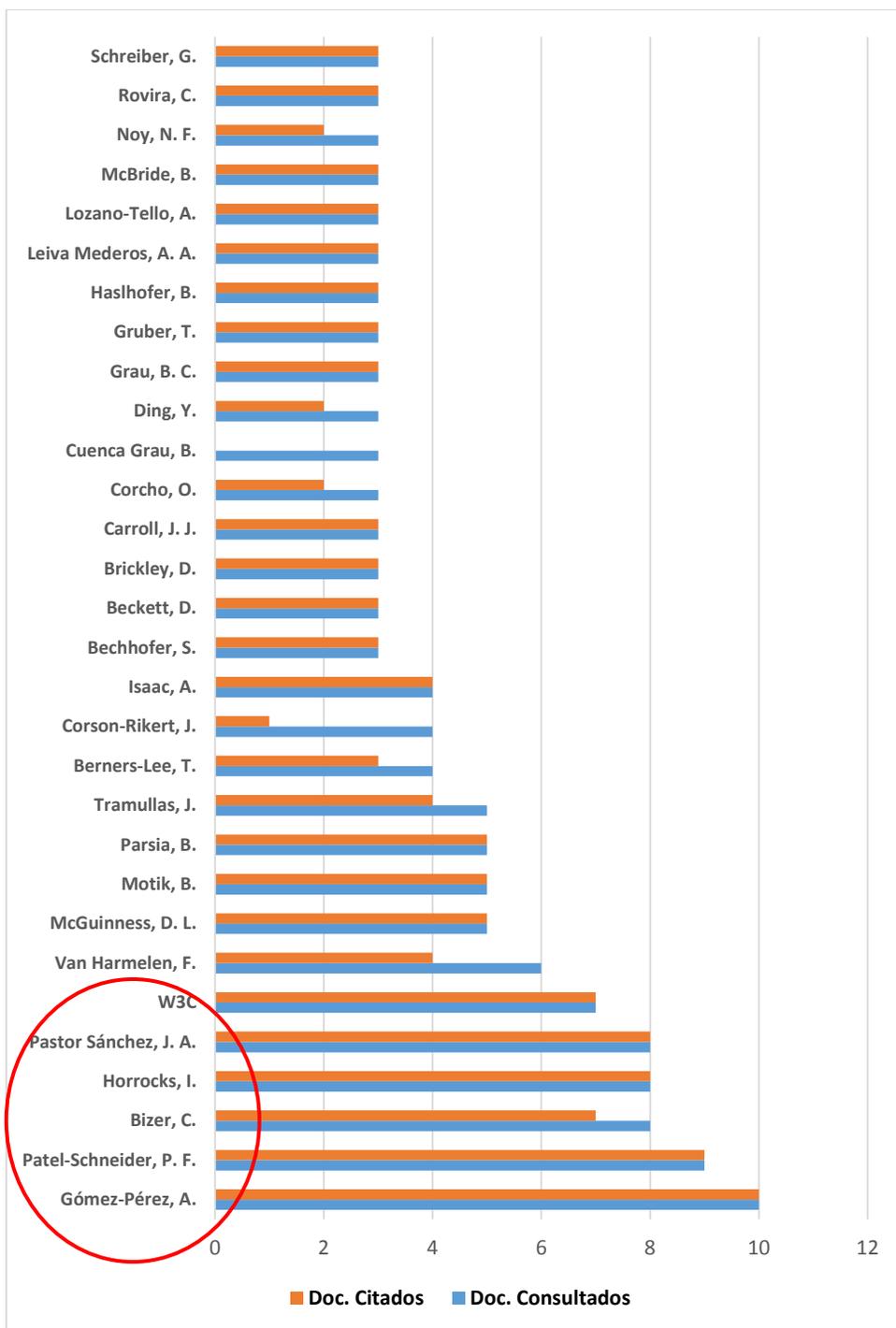


Fig. 4.5. Relación entre la cantidad de documentos consultados y los citados de los principales autores trabajados (elaboración propia)

Pueden observarse varios autores que en función de la cantidad de documentos consultados y citados, pueden considerarse como los más productivos en la elaboración de la tesis:

- Asunción Gómez-Pérez: Es Vicerrectora de Investigación, Innovación y Doctorado. Es Catedrática en la Universidad Politécnica de Madrid (UPM). Fue la directora del Departamento de Inteligencia Artificial (2008-2016), directora del Ontology Engineering Group (<http://www.oeg-upm.net/>) en la UPM (desde 1995) y Directora Académica del Máster Universitario en Inteligencia Artificial (desde 2009) y del Doctorado en Inteligencia Artificial en la UPM (desde 2009). Entre sus áreas de investigación y visibilidad internacional se incluyen la ingeniería ontológica, la web semántica, los datos enlazados, el multilingüismo en datos y la gestión de conocimientos. Ha participado en la creación a nivel mundial del área de ontologías y Web semántica desde sus comienzos, en la década de los noventa. Una muestra de su proyección internacional se presenta en el estudio de T. Chambers, S. Mijojevic, Y. Ding. Female Semantic Web Researchers: Does Collaboration with Male Researchers Influence their Network Status, en donde se le identifica como una de las tres mujeres más exitosas e influyentes en temas de Web semántica a nivel mundial. (<http://www.dia.fi.upm.es/?q=es/node/63>)
Dicha autora se distingue por un alto nivel de colaboración con otros autores (Ver Fig.4.7). Por lo general aparece en un 2 o tercer nivel de presentación de autores en los documentos publicados.
- Peter F. Patel-Schneider: Actualmente funge como Nuance Communications del W3C (<https://www.linkedin.com/in/peter-f-patel-schneider-5a65b848>) Recibió su Ph. D. en Ciencias de la Computación de la Universidad de Toronto en 1987. De 1983 a 1988 fue en el grupo de investigación en IA en Fairchild y Schlumberger. Se unió al Departamento de Investigación de Principios de IA en el AT & T Bell Laboratories, y luego fue a AT & T Labs - Investigación cuando AT & T se separó en 1995. En 1997 se reunió con los Laboratorios Bell de Investigación y permaneció allí hasta 2012. Ha impartido cursos tanto en la Universidad de Toronto y la Universidad de Rutgers. Sus intereses de investigación refieren las propiedades y uso de lógicas descriptivas,

en particular el W3C OWL Web Ontology Language. Ha diseñado e implementado grandes sectores de CLASSIC, un sistema de representación del conocimiento basado en la lógica Descripción. Se diseñó e implementó DLP, una cámara de fermentación optimizada fuertemente para lógicas descriptivas y expresivas lógica modal proposicional. Ha realizado una extensa evaluación empírica de DLP y otros experimentadores para lógicas descriptivas y la lógica modal proposicional. Desarrolló gran parte de OWL, y su predecesor DAML + OIL, así como SWRL, la semántica del lenguaje del Web Regla, y OWL 2, la reciente revisión de OWL. Peter ha estado trabajando recientemente en la extracción de información semántica a partir de fuentes de datos, permitiendo que los datos que deben integrarse más fácilmente en la Web Semántica o la Web de Datos. Ha sido determinar cómo utilizar el cálculo paralelo para el procesamiento semántico más eficaz de grandes cantidades de datos. Pedro también ha estado investigando cómo representar y razonar con los servicios, en particular para la detección de servicios semi-automática.

Dicho autor se distingue por un alto nivel de colaboración con otros autores (Ver fig. 4.7). Por lo general aparece en un segundo o tercer nivel de presentación de autores en los documentos publicados.

- Christian Bizer: Es Director del Instituto de Ciencias de la Computación e Informática de negocios y profesor de la Universidad de Mannheim. Explora cuestiones técnicas y empíricos sobre el desarrollo de entornos de información global y descentralizada. Su investigación actual se centra la evolución de la World Wide Web desde un medio para la publicación de documentos en un espacio de datos global. Bizer ha inicializado el esfuerzo de la comunidad W3C Linking Open Data que está interconexión gran número de fuentes de datos en la Web. Fue co-fundador del proyecto DBpedia que se deriva de una amplia base de conocimientos de Wikipedia. También inicializa el proyecto Web Data Commons que supervisa el despliegue de RDFa, de microdatos, microformatos y datos en la Web mediante el análisis de grandes

arrastra Web. Otros resultados de su trabajo incluyen el modelo de datos gráficos con nombre que se adoptó en las recomendaciones del W3C RDF y SPARQL; el lenguaje de mapeo D2RQ que es ampliamente utilizado para la publicación de las bases de datos relacionales a la Web de Datos; la Seda - Comparación de identidades Marco, y el punto de referencia de Berlín SPARQL para medir el rendimiento de las tiendas de RDF. Cristiano Bizer tenga un nombramiento como profesor de tiempo completo en la Universidad de Mannheim, donde dirige los sistemas basados en la Web del Grupo de Enfoque de los datos y el Grupo de Ciencia Web. Antes de trasladarse a Mannheim, dirigió el grupo de sistemas basados en la Web en la Universidad Libre de Berlín. (<http://dws.informatik.uni-mannheim.de/en/people/professors/prof-dr-christian-bizer/>)

Dicho autor, generalmente comparte autoría con otros autores (ver fig.4.7).

- Ian Robert Horrocks: Es profesor de Ciencias de la Computación en la Universidad de Oxford en el Reino Unido y un miembro de Oriel College, Oxford. Sus intereses de investigación incluyen la representación del conocimiento, ontologías y lenguajes de ontologías, modelos y lógicas de descripción, el razonamiento automatizado, implementación y optimización de los sistemas de razonamiento, y aplicaciones en áreas tales como e-Science y la Web Semántica. (<http://www.cs.ox.ac.uk/ian.horrocks/>)

Según valoraciones realizadas por el grupo de trabajo de la Wikipedia, su trabajo en el razonamiento cuadro de lógicas descriptivas muy expresivos ha formado la base de la mayoría de los sistemas de razonamiento Descripción de lógica en uso hoy en día. (https://en.wikipedia.org/wiki/Ian_Horrocks)

Dicho autor se distingue por un alto nivel de colaboración con otros autores. (Ver fig. 4.7)

- Juan Antonio Pastor Sánchez: Es profesor contratado doctor permanente del Departamento de Información y Documentación de la Universidad de

Murcia. Ha desempeñado su labor desde el curso 2002/2003, principalmente en las titulaciones de Licenciado en Periodismo y Licenciado en Publicidad y Relaciones Públicas en la Facultad de Comunicación y Documentación, aplicando en el campo docente su experiencia profesional de más de 15 años en el desarrollo de servicios y productos de información, sistemas de información y bases de datos para la Web, creación de entornos interactivos de enseñanza y el diseño web realizado con criterios de Accesibilidad, Usabilidad y Arquitectura de la Información.

Las actividades de investigación desarrollada han girado alrededor de las varias líneas de trabajo: Teoría y modelos de hipertexto; Automatización en la Gestión de Tesoros; Modelos tecnológicos para la Gestión de Información y la Organización del Conocimiento; Tecnologías y especificaciones de la Web Semántica; Accesibilidad, Usabilidad y Arquitectura de la Información para la Web (<http://www.um.es/docencia/pastor/bio/>).

Dicho autor, a pesar de su productividad, para los registros consultados, los niveles de colaboración son bajos. Generalmente aparece como único autor y eventualmente comparte autoría con otros autores que no parecen ser sostenidas.

- W3C: World Wide Web Consortium: Es una comunidad internacional donde las organizaciones miembros⁶⁶, personal⁶⁷ a tiempo completo y el público en general trabajan conjuntamente para desarrollar estándares Web. Liderado por el inventor de la Web Tim Berners-Lee y el Director Ejecutivo (CEO) Jeffrey Jaffe. La misión del W3C es guiar la Web hacia su máximo potencial (<http://www.w3c.es/Consortio/>).

En la tabla quedan igualmente evidenciado el uso de autores clásicos como Berners-Lee⁶⁸.

⁶⁶ Organizaciones Miembros de la W3C: <https://www.w3.org/Consortium/Member/List>

⁶⁷ Personal a tiempo completo del W3C: <https://www.w3.org/People/>

⁶⁸ Página personal de Tim Berners Lee: <https://www.w3.org/People/Berners-Lee/>

Otro de los elementos que distingue el trabajo autoral es la colaboración que suele establecerse entre ellos para el desarrollo de proyectos y publicaciones. En las figuras que se presentan a continuación (ver fig. 4.6 y 4.7) se muestra la colaboración entre los autores de las fuentes consultadas en la tesis.

Para el análisis de colaboración autoral, se descartaron, inicialmente, los datos de 88 autores que no que no compartían mención de responsabilidad en ninguno de los registros consultados. Dentro de ellos se incluyen los autores institucionales como las universidades y centros u organizaciones de investigación.

En la matriz visualizada en la fig.4.6, se muestran 511 relaciones de colaboración autoral. Dada la complejidad de su lectura se realiza una poda para visualizar las relaciones establecidas entre aquellos autores que exhiben dos o más relaciones de colaboración (ver fig. 4.7), en aras de obtener una red que sea comprensible y que muestre las relaciones de colaboración entre los autores utilizados en la elaboración del informe.

La figura 4.7 muestra, a través del tamaño del nodo, la cantidad de relaciones establecidas con otros autores (valor de intermediación) y, con el grosor de la línea, la fortaleza de dichas relaciones.

Los autores que se distinguen por el grado de colaboración son: Van Harmelen, Patel- Schneider y Horrocks

Se diferencian, además, Asunción Gómez Pérez y Christian Bizer.

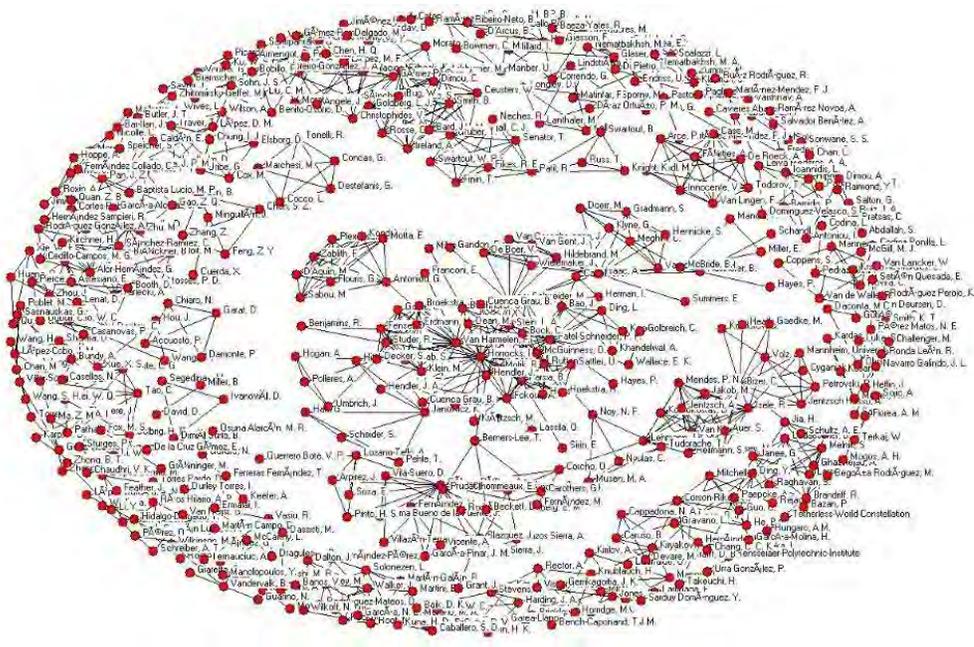


Fig. 4.6. Colaboración autorral de la bibliografía consultada (elaboración propia)

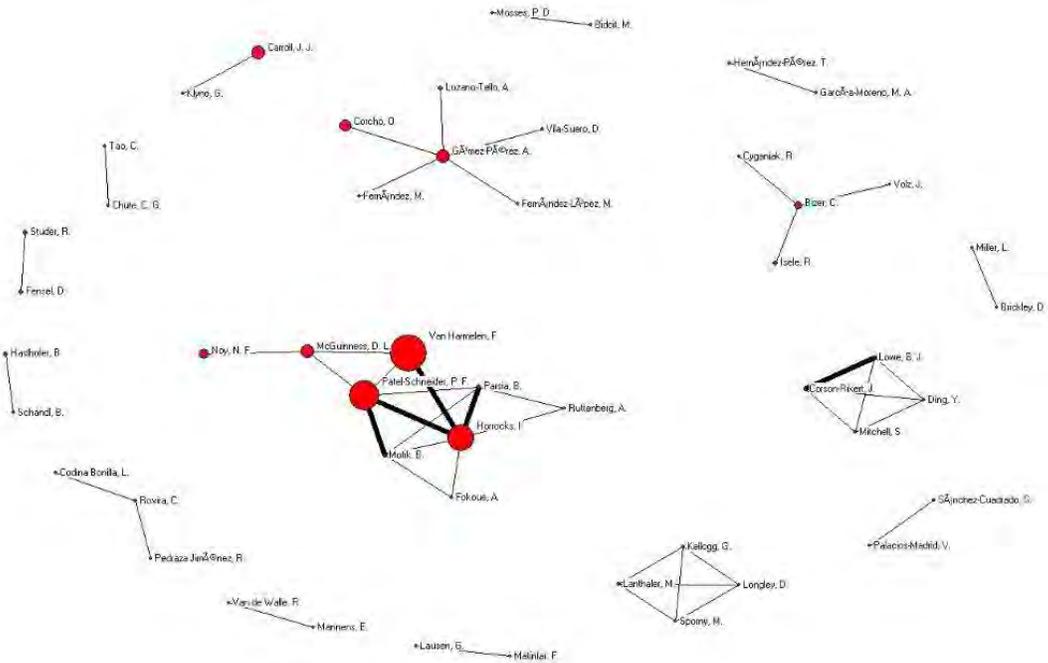


Fig. 4.7. Colaboración autorral de la bibliografía consultada: Matriz podada (elaboración propia)

4.10. Referencias bibliográficas

- ANDREU-VALL, M. & MARCOS, M.-C. 2012. Evaluación de sitios web multilingües: metodología y herramienta heurística. *El profesional de la información*, 21, 254-260.
- ARENCIBIA JORGE, R. 2010. *La colaboración científica entre Cuba y España en el Web of Science 1999-2003*. Tesis doctoral, Universidad de Granada.
- BOLAÑOS-PIZARRO, M., VIDAL-INFERRER, A., NAVARRO-MOLINA, C., VALDERRAMA-ZURIÁN, J. C. & ALEIXANDRE-BENAVENT, R. 2007. Usabilidad: concepto y aplicaciones en las páginas web médicas. *papeles médicos*, 16, 14-21.
- FERRÀS FONT, D., GARCÍA ALMIÑANA, J. & AINOZA FARLED, J. 2007. *Estudio y evaluación de sistemas gestores de contenidos web*. Master thesis, Universitat Politècnica de Catalunya.
- HASSAN MONTERO, Y. & MARTÍN FERNÁNDEZ, F. J. 2003. Guía de evaluación heurística de sitios web. *No Solo Usabilidad*.
- HERNÁNDEZ SAMPIERI, R., FERNÁNDEZ COLLADO, C. & BAPTISTA LUCIO, M. D. P. 2010. *Metodología de la Investigación*, The McGraw-Hill Companies.
- KERLINGER, L. 2006. *Metodología de la investigación*, McGraw-Hill Interamericana.
- TRAMULLAS. 2006. Intranet: Modelo para el Desarrollo de Redes Corporativas. [Online]. Available: <http://www.tramullas.com/pdf/intranet.pdf>. [Accessed 15 de noviembre de 2015].
- URÍAS, G. 2009. *Metodología de la Investigación*, Santa Clara, Universidad Central "Marta Abreu" de las Villas.
- VIZCAYA ALONSO, D. (ed.) 1997. *Información: procesamiento de contenido*, Rosario: Nuevo Paradigma.

CAPÍTULO V

MARCO APLICADO DE LA INVESTIGACIÓN: LA UNIVERSIDAD DE LA HABANA

CAPÍTULO 5: Marco aplicado de la investigación: la Universidad de La Habana.

El presente capítulo tiene como objetivo describir la Universidad de La Habana, los sistemas de gestión automatizados y productos de gran importancia para la Universidad construidos con CMS.

5.1. Introducción

La Sociedad de la Información está generando profundas transformaciones en nuestra forma de vivir, de percibir las necesidades y ofrecer soluciones, algo que repercute en todas las esferas sociales, desde la economía hasta la política. El sector de la educación no está exento de ello. En este nuevo modelo social, centrado en la generación de conocimiento y procesamiento de la información, el papel de las universidades es clave para el desarrollo.

Las NTIC al ser puestas función de la instrucción universitaria comienzan a desencadenar procesos de aprendizaje con la finalidad de orientar al estudiante hacia la creación de su propio conocimiento, a partir del conjunto de recursos e informaciones disponibles. Para acometer esta nueva tarea el profesor necesita trabajar conjuntamente con diversos especialistas (informáticos, pedagogos, profesionales de la información, diseñadores, etc.); lo que exige del trabajo cooperativo, pues la simple presencia de una computadora conectada a Internet o en red no garantiza un adecuado aprovechamiento de las tecnologías, pese a que sus posibilidades sean infinitas. Las NTIC no generan automáticamente nuevos saberes. La herramienta ha de estar al servicio de lo educativo y no a la inversa.

El comportamiento, manifestación y tendencias de las NTIC en el contexto universitario son fenómenos de vital importancia por las repercusiones que, directa o indirectamente, aporta a la dinámica de la sociedad en todas sus esferas.

El impacto de los CMS no se puede obviar. La rebaja en los costos del proceso de elaboración e implementación de los productos electrónicos para la Web, la disminución del tiempo invertido, la ramificación de la cantidad de personas que pueden trabajar en la creación y actualización (multiplicando la fuerza de trabajo) y el aumento de las potencialidades desarrolladas para el producto final, son elementos trascendentales que justifican el uso de los CMS.

5.2. Universidad de La Habana: Caracterización

La Universidad de La Habana fue fundada en 1728. Se distingue por su implicación directa en los procesos sociales y revolucionarios que se han desarrollado en el país a lo largo de su casi tricentenaria historia y por ser sede de acontecimientos trascendentales. "Es referente cultural y de identidad para Cuba y el mundo, y asume con excelencia y alto compromiso político los retos del desarrollo de la educación, la ciencia y la cultura" (UniversidadDeLaHabana, 2010). Tiene como misión garantizar los requerimientos de generación de conocimiento científico y de formación integral y continua de profesionales revolucionarios, altamente calificados, con capacidad científica y política que permitan la ejecución de proyectos teóricos y prácticos vinculados a la construcción de la sociedad socialista y el desarrollo sostenible del país.

Cuenta para ello con su acervo de conocimientos multifacéticos, el liderazgo académico que le confieren casi tres siglos de creación e impacto en la formación de la nación cubana y un reconocido prestigio nacional e internacional en las ciencias naturales, exactas, económicas, sociales y humanísticas. Forma profesionales en 32 carreras universitarias de ciencias naturales y exactas, sociales y económicas, bajo diferentes modalidades de estudio, tal como presenta la tabla 5.1 (Ver además las figuras de la 5.5 a la 5.35).

CURSO REGULAR DIURNO (CRD)	TOTAL	1ero	2do	3ero	4to	5to
Totales	7320	1948	1449	1347	1312	1264
1. Biología	315	86	73	57	46	53
2. Bioquímica	272	79	61	51	41	40
3. Ciencias Alimentarias	234	53	54	26	52	49
4. Ciencias de la Computación	390	148	83	48	62	49
5. Ciencias de la Información	230	40	43	38	66	43
6. Ciencias Farmacéuticas	291	63	75	64	62	27
7. Comunicación Social	233	44	39	43	56	51
8. Contabilidad y Finanzas	653	170	128	119	129	107
9. Derecho	629	125	103	115	140	146
10. Economía	587	159	111	122	95	100
11. Filosofía Marxista-Leninista	126	35	14	19	28	30
12. Física	120	50	26	21	9	14
13. Geografía	272	89	64	58	32	29
14. Historia	193	40	36	40	43	34
15. Historia del Arte	258	49	35	70	60	44
16. Ingeniería Física	17	0	0	0	9	8
17. Lengua Alemana (con segunda Lengua Extranjera)	106	25	22	19	14	26
Lengua Alemana (preparatoria)	27	27	0	0	0	0
18. Lengua Española para Estudiantes no Hispanohablantes	23	14	8	0	1	0
19. Lengua Francesa (con segunda Lengua Extranjera)	91	21	13	16	23	18
Lengua Francesa (preparatoria)	25	25	0	0	0	0
20. Lengua Inglesa (con segunda Lengua Extranjera)	121	30	20	22	23	26
Lengua Inglesa (Preparatoria)	38	38	0	0	0	0
21. Lengua Rusa (con segunda Lengua Extranjera)	38	10	13	9	0	6
Lengua Rusa (preparatoria)	21	21	0	0	0	0
22. Letras	239	66	54	49	35	35
23. Matemática	141	49	40	31	21	0
24. Microbiología	213	54	39	69	22	29
25. Periodismo	241	30	31	27	60	93
26. Psicología	447	116	97	92	72	70
27. Química	246	73	61	39	44	29
28. Sociología	180	40	36	30	24	50
29. Turismo	303	79	70	53	43	58

Fig. 5.1. Carreras que se estudian en el curso regular diurno (Cobreiro, 2015)

CURSO POR ENCUENTROS (CPE)	TOTAL	1ero	2do	3ero	4to	5to	6to	Sin Año
Totales	3336	354	613	475	421	467	897	2
1. Bibliotecología y Ciencia de la información	102	23	47	20	1	9	0	2
2. Ciencias Alimentarias	83	8	19	18	11	6	21	0
3. Ciencias Farmacéuticas	92	10	16	17	17	8	24	0
4. Comunicación Social	145	4	0	19	53	69	0	0
5. Derecho	403	80	63	53	89	118	0	0
6. Estudios Socioculturales	303	0	92	47	32	24	108	0
7. Preservación y Gestión del Patrimonio Histórico	30	24	3	0	2	1	0	0
8. Turismo	398	0	0	0	8	47	343	0
9. Historia	107	29	39	22	17			

Fig. 5.2. Carreras que se estudian por encuentro (Cobreiro, 2015)

CONTINUIDAD DE ESTUDIOS (CE)	TOTAL	5to	6to	Sin Año
Totales	1805	10	1779	16
1. Bibliotecología y Ciencia de la información	19	0	19	0
2. Comunicación Social	279	0	279	0
3. Contabilidad y Finanzas	440	0	440	0
4. Derecho	450	0	434	16
5. Historia	6	0	6	0
6. Psicología	605	10	595	0
7. Sociología	6	0	6	0

Fig. 5.3. Carreras que se estudian en continuidad de estudio (Cobreiro, 2015)

EDUCACIÓN A DISTANCIA (ED)	TOTAL UH
TOTAL UH	1830
1. Ciencias de la Información	15
2. Derecho	921
3. Estudios socioculturales	482
4. Historia	60
5. Economía	11
6. Contabilidad y finanzas	333
7. Turismo	8

Fig. 5.4. Carreras que se estudian a distancia (Cobreiro, 2015)

La estructura universitaria comprende 17 facultades⁶⁸), 6 centros de estudios⁶⁹ no adscritos a facultades, 5 entidades de ciencia, tecnología e innovación, 2

⁶⁸ Además de las 16 facultades representadas en la figura 5.6, también es considerada como facultad el IFAL.

departamentos centrales y 6 residencias estudiantiles. Tiene tres unidades presupuestadas adscritas: la de Aseguramiento (UPA), el Jardín Botánico Nacional (JBN) y el Instituto de Farmacia y Alimentos (IFAL) (ver figura 5.6 y Anexo 4). Dadas sus dimensiones estructurales y constructivas, el campus universitario capitalino está desplegado por toda la ciudad (ver fig.5.5)

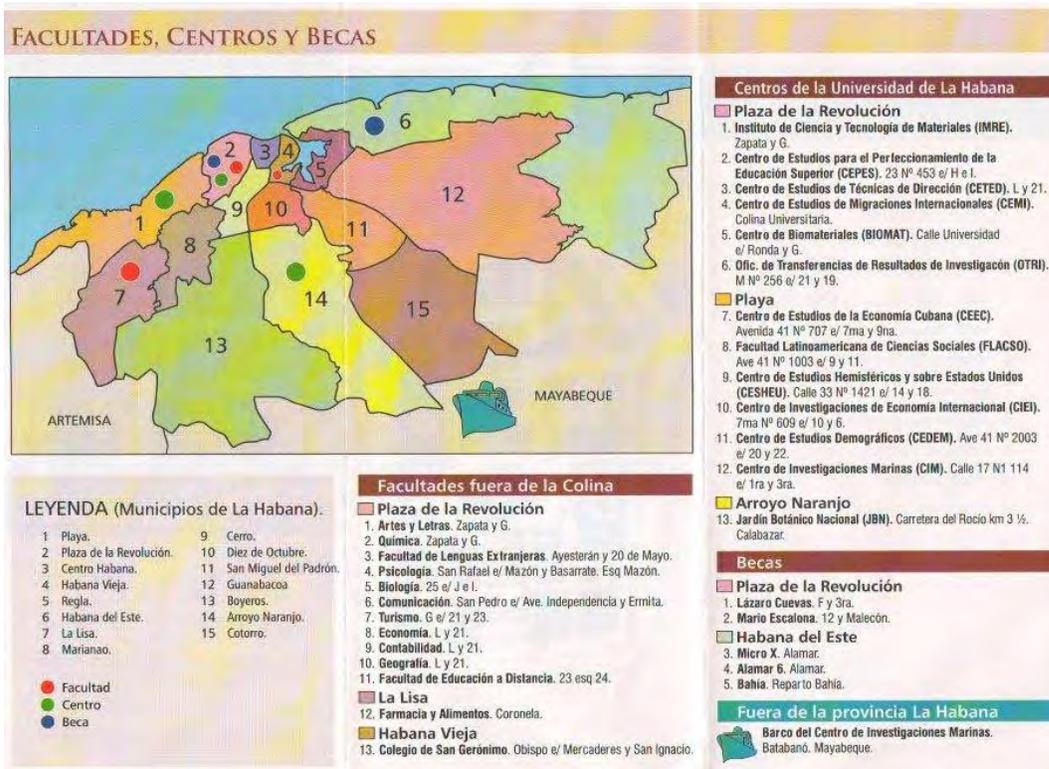


Fig. 5.5. Distribución geográfica de las diferentes estructuras de la UH

⁶⁹ Además de los 5 centros representados en la figura 5.6, también es considerada parte de este grupo el Instituto Confucio.

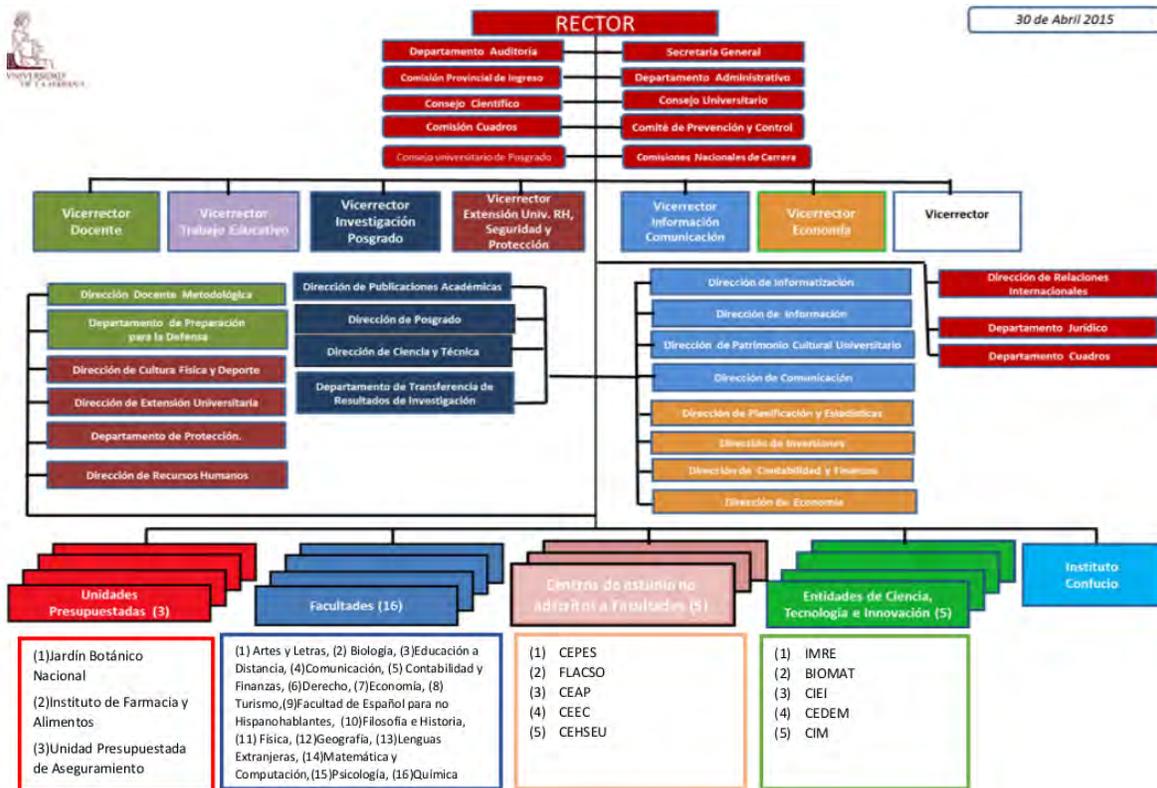


Fig. 5.6. Organigrama de la Universidad de La Habana

Puede advertirse que la UH pese una estructura compleja, no solo desde el punto de vista estructural, también temático y profesional.

5.3. Intranets en la UH

Las universidades cubanas han estado inmersas en el desarrollo de diversos proyectos que implican la creación de productos digitales en función de la docencia, la producción científica, la gestión y la comunicación (tanto interna como externa. Díganse pues: páginas web institucionales, intranets, bibliotecas digitales y virtuales, sitios Web de asignaturas, repositorios, productos multimedia, desarrollo de software educativos). La política aplicada a la gestión de las tecnologías es decisiva para el desarrollo estratégico de cualquier organización y debe ser coherente con las condiciones económicas y el desarrollo científico del centro.

Los beneficios de las aplicaciones de los CMS en el desarrollo de productos Web en la educación superior son multidimensionales, intangibles y prácticamente imposibles de contabilizar en términos económicos.

En la Universidad de la Habana, internet ha tenido no solo fines investigativos, también se implementan y usan las páginas Web institucionales para difundir informaciones sobre sus capacidades, funciones, lineamientos generales, contactos, etc., es decir, a los elementos que definen su imagen corporativa. La intranet ha facilitado la promoción, difusión, y uso de la información, así como la prestación de una serie de servicios que van desde una simple búsqueda hasta la adquisición de un producto o información de cualquier tipo. Van estrechamente vinculadas a la gestión estratégica de la información dentro de la UH y están llamadas a convertirse en el principal y más importante sistema de manejo de la información en línea de la universidad.

Durante el curso 2006-2007 se comenzó a implementar Plone en la UH para el sistema de las Intranet de algunas facultades. Su uso se extendió durante el curso 2007-2008, pero con el cambio de las personas que dirigían metodológicamente el trabajo a nivel de la Universidad, se tomó la decisión de migrar a un nuevo CMS. Se argumentaron diversas complejidades e inconvenientes con Plone asociadas esencialmente con:

- El lenguaje de programación, se tornaba difícil el desarrollo de módulos para adaptarlos a las necesidades de la Universidad.
- Dificultades con implementación de otro CMS en el mismo servidor en el que corría Plone

Ello llevó al auge de un movimiento para el estudio de los CMS en la Universidad. Aunque la UH no impuso el uso de ninguno en particular, la mayoría de las facultades apostaron por el trabajo entre Joomla y Drupal. Esta puede ser la posible explicación del nivel de popularidad de dichos CMS es este contexto.

En estudio exploratorio realizado en 2009 por la autora (Rosell León, 2009), el 50% de las intranets existentes, estaban implementadas en algún tipo de CMS. Se identificaron un total de 5 CMS, de los cuales 3 son de tipo software libre. (Ver Fig.5.7)

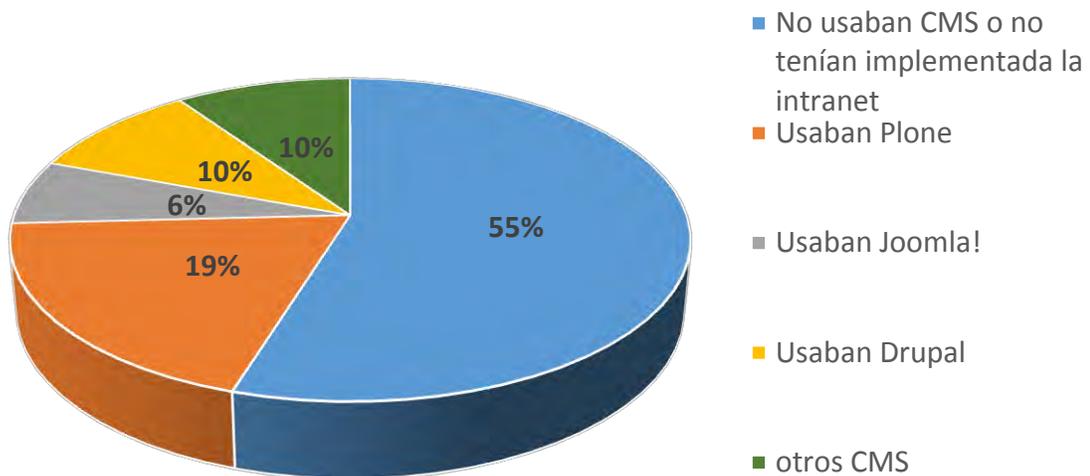


Fig. 5.7. CMS identificados en la UH (intranets) en estudio exploratorio de 2009 (elaboración propia)

En aquel momento (septiembre, 2009), la UH contaba con 18 facultades de las cuales 11 contaban con intranet o/y página web institucional construida con algún tipo de CMS:

- 6 facultades usaron Plone
- 3 facultades usaron Joomla!
- 2 facultades usaron Drupal

Pasado más de un quinquenio del estudio exploratorio, los CMS y el impacto generado en su uso puede verse desde diversas perspectivas. En la investigación se toma como punto de referencia a las Intranet pues son el factor común entre las diferentes estructuras de la Universidad. Se pretende presentar las construidas desde las distintas estructuras de la UH, para contabilizar el uso de los CMS y las razones de que han llevado a ello, como caracterización del marco aplicado de la investigación.

Actualmente, en la UH de 31 estructuras 26 cuentan con una intranet y solo 5 aún no la implementan de manera oficial (Ver Fig.5.8), pues se tiene conocimiento que las tienen en proyecto. Las estructuras sin intranet publicadas son: CEHSEU, Instituto Confucio, IMRE, Jardín Botánico y la UPA.

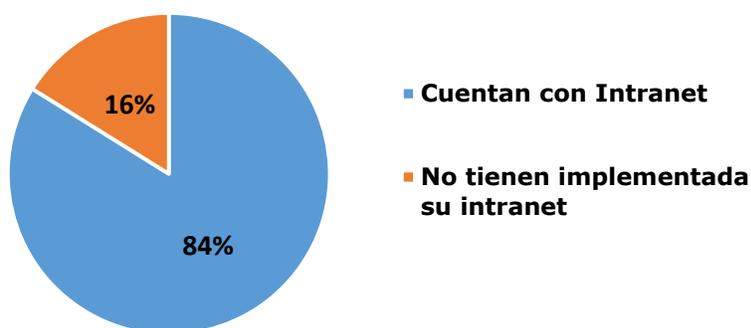


Fig. 5.8. Existencia de Intranets en la UH, enero/ 2016 (elaboración propia)

Las intranets son construidas y gestionadas desde las áreas, aunque desde la Dirección de Informatización y la Dirección de Comunicación de la universidad se monitorea constantemente. Por esta razón, aunque casi todas las áreas cuentan con su intranet, no existe uniformidad en el diseño o el software que se utiliza.

5.3.1. Universidad de La Habana

La nueva versión de la intranet de la UH (<http://intranet.uh.cu/>), fue publicada en el año 2010, por el Grupo "Producciones IRIS" con el CMS Drupal. (Ver Fig.5.40). El cambio fundamental con respecto al diseño de la intranet anterior⁷¹, radica en el sistema de navegación basado en la gestión por procesos. (ver fig. 5.9 y 5.10)

⁷¹ La intranet anterior se basaba en la estructura administrativa de la UH lo que hacía compleja la navegación y la recuperación de información.



Fig. 5.9. Página de inicio de la Intranet UH (enero/2009)



Fig. 5.10. Página de inicio de la Intranet UH (enero/2016)

En la Intranet UH convergen un conjunto de servicios e información clave sobre procesos sustantivos: Estudios, Investigaciones, Actividades de

Extensión Universitaria (extensión, actividades, eventos) y Gestión Institucional.

Es un espacio pensado para contribuir al desempeño organizacional más eficiente y oportuno. A decir de sus diseñadores y responsables de desarrollo, su diseño conceptual se encuentra en la primera etapa de implementación.

Se caracteriza por:

- La información es presentada por procesos: esto significa que es posible encontrar información sin conocer quién la produce y/o dónde se ubica en la estructura institucional.
- Los contenidos de las secciones son administrados por las áreas centrales a las que corresponden.
- Posee un Consejo Editorial, que se reúne semanalmente para impulsar su desarrollo y establecer las prioridades informativas y comunicativas,

En la fig. 5.11, se señalan los principales menús de navegación y algunas áreas importantes de la página principal.



Fig. 5.11. Diferentes áreas de la Página de inicio de la Intranet UH

5.3.2. Facultad de Artes y Letras

La Facultad de Artes y Letras prepara a sus estudiantes para la práctica, el disfrute, la investigación y la crítica del arte, la lengua y la literatura, la construcción de sus variadas y diferenciadas historias y la enseñanza de tan complejas cuestiones, así como para el diagnóstico de problemas sociales, y

la intervención por la vía del trabajo cultural. Se estudian las carreras de Licenciatura en Historia del Arte y Licenciatura en Letras.

Las líneas principales de investigación se orientan hacia el arte, la literatura y la lengua, particularmente en el entorno americano, caribeño, antillano y cubano, tanto en sus vertientes históricas, reconstructivas y teóricas, como en el abordaje crítico de autores, artistas y fenómenos, para darlos a conocer y promoverlos. Ello se lleva a cabo en permanente diálogo con las figuras y corrientes internacionales.

La Intranet de la Facultad de Artes y Letras (<https://portal.fayl.uh.cu/>) fue publicada en 2015 (Ver fig.5.12), anteriormente no contaba con una. Es un sitio implementado en Drupal.

Su arquitectura sigue un patrón similar al de la intranet de la UH.

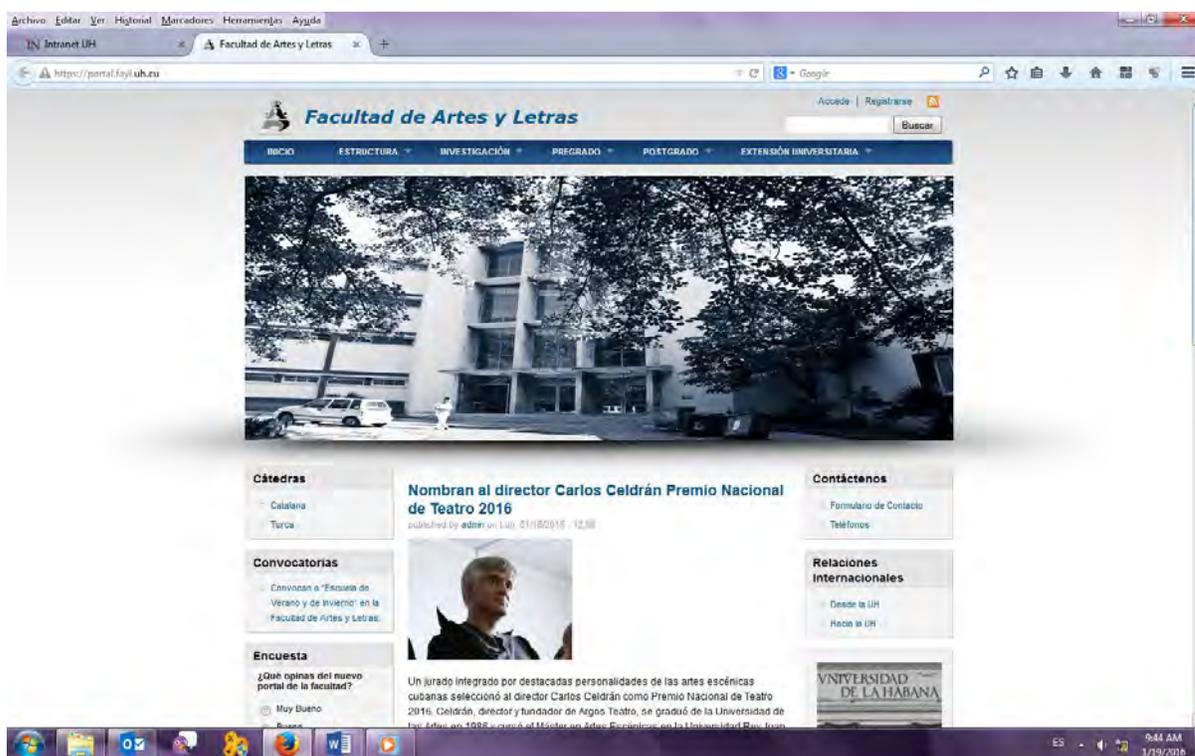


Fig.5.12. Página de inicio de la Intranet de la Facultad de Artes y Letras

5.3.3. Facultad de Biología

Surge como Facultad en el año 1976. Es el centro rector de la enseñanza de las Ciencias Biológicas en Cuba. Su objetivo es garantizar “la formación de biólogos, microbiólogos y bioquímicos, así como la superación postgraduada de profesionales de centros de investigación y de producción y servicios de todo el país, con un alto grado de desarrollo científico-técnico y una **adecuada actualización de conocimientos.**”(IRIS, 2011). Se estudian las carreras de Licenciatura en Biología, Licenciatura en Bioquímica y Licenciatura en Microbiología.

Como centro rector de la enseñanza de las Ciencias Biológicas en Cuba, entre sus objetivos está contribuir con el desarrollo científico, económico y social del país, a través de la actividad docente y científico-investigativa; garantizar la formación de biólogos, microbiólogos y bioquímicos, así como la superación postgraduada de profesionales de centros de investigación y de producción y servicios.

La nueva versión de la intranet (<http://intranet.fbio.uh.cu/>) fue publicada en 2015, e implementada en Drupal. (Ver Fig. 5.13)



Fig. 5.13. Página de inicio de la intranet de la facultad de Biología

5.3.4. Facultad de Comunicación

La Facultad de Comunicación lleva cuatro décadas en la formación de profesionales. En ella se estudian las carreras de Periodismo, Comunicación Social y Ciencias de la Información. La docencia y la investigación se realizan en un espacio de liderazgo para desarrollar y promover la integración interdisciplinaria de la comunicación y la información, sectores claves para el avance actual y futuro de nuestra sociedad. Asimismo la Facultad interactúa con las demandas profesionales y sociales en los diferentes ámbitos de desempeño a nivel comunitario, territorial y nacional.

La nueva versión de intranet construida en Drupal fue publicada en 2014 (<http://intranet.fcom.uh.cu/>) (Ver fig.5.14)



Fig. 5.14. Página de inicio de la intranet de la Facultad de Comunicación

5.3.5. Facultad de Contabilidad

La Facultad de Contabilidad y Finanzas garantiza la formación y la capacitación de profesionales vinculados a las ciencias contables, financieras y administrativas, con el objetivo de que desarrollen conocimientos y

habilidades que les permitan tener un desempeño laboral competente, y de que científicamente se sientan capacitados para actuar de manera independiente y creadora, acorde a las necesidades del desarrollo del país; además para que posean un alto compromiso con el proyecto social cubano y sean capaces de contribuir con las transformaciones que se deriven del desarrollo de la sociedad cubana dentro del sistema de valores del socialismo.

En la Facultad de Contabilidad y Finanzas se estudia las carreras de Licenciatura en Contabilidad y Finanzas, Licenciatura en Geografía y Licenciatura en Economía.

La Intranet de la Facultad de Contabilidad y Finanzas (<http://intranet.fcf.uh.cu/>) fue publicada en 2014 (Ver fig.5.15). Es un sitio implementado en Joomla.



Fig.5.15. Página de inicio de la intranet de la Facultad de Contabilidad y Finanzas

5.3.6. Facultad de Derecho

La Facultad de Derecho, continuadora de la Facultad de Leyes y Jurisprudencia, creada en 1728 al constituirse la Real y Pontificia

Universidad de San Jerónimo de La Habana, forma al jurista no como un simple concedor e intérprete más o menos agudo del ordenamiento jurídico positivo, sino que provee de una formación científica, doctrinal y técnica, en directa y declarada reacción antinormativista, que hacen al profesional capaz para el ejercicio de la abogacía, de la actuación fiscal y jurisdiccional y para la asesoría jurídica.

En la Facultad de Derecho se estudia la carrera de Licenciatura en Derecho. Tiene su intranet (<http://intranet.lex.uh.cu/>) implementada en Drupal y fue publicada su nueva versión en 2011. (Ver Fig. 5.16)

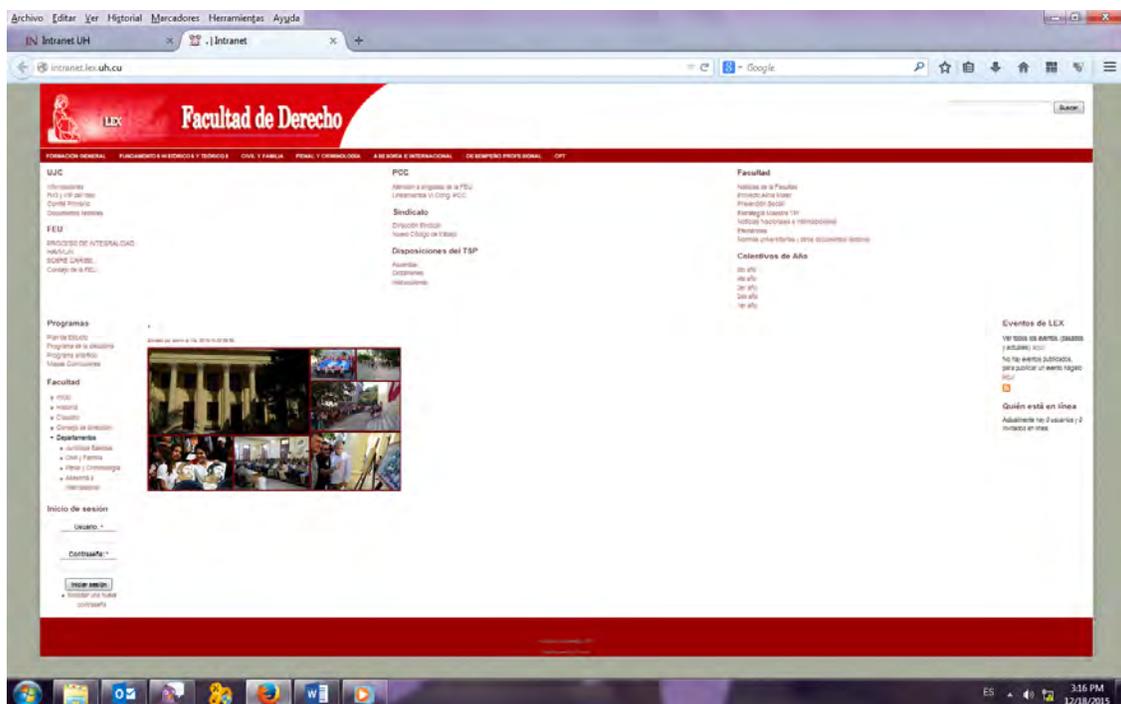


Fig. 5.16. Página de inicio de la intranet de la Facultad de Derecho

5.3.7. Facultad de Economía

La Facultad de Economía de la Universidad de La Habana, con más de cuarenta años de fundada, es la institución dedicada a la formación de especialistas en las ciencias económicas y constituye el centro rector para la enseñanza de sus especialidades en el país.

La Facultad tiene como misión formar un economista integral, calificado, culto y comprometido con los principios de la Revolución, que con la

complementación de la enseñanza de postgrado y una intensa labor de investigación, contribuya al perfeccionamiento del modelo socialista cubano, desde una incondicional vocación social y como visión un plan de estudios que es referente latinoamericano para la formación de economistas y ofertamos programas de postgrado y servicios científico - técnicos con proyección nacional e internacional.

En la Facultad de Economía se estudia la carrera de Licenciatura en Economía

La Intranet de la Facultad de Economía (<http://fec.uh.cu/>) fue publicada en 2012 (Ver fig.5.17). Es un sitio implementado en Drupal.



Fig. 5.17. Página de inicio de la intranet de la Facultad de Economía

5.3.8. Facultad de Turismo

La Facultad fue creada con el objetivo de contribuir al desarrollo del turismo en Cuba, mediante la realización de investigaciones, servicios científico-técnicos, consultorías y asesorías; así como con la realización de actividades

docentes dirigidas tanto a la formación profesional como a la superación de postgrado de los directivos y especialistas que demanda el sector .

La Intranet de la Facultad de Turismo ([http://intra.ftur.uh.cu./](http://intra.ftur.uh.cu/)) fue publicada en 2016 (Ver fig.5.18), anteriormente no contaba con una. Es un sitio implementado en Drupal.

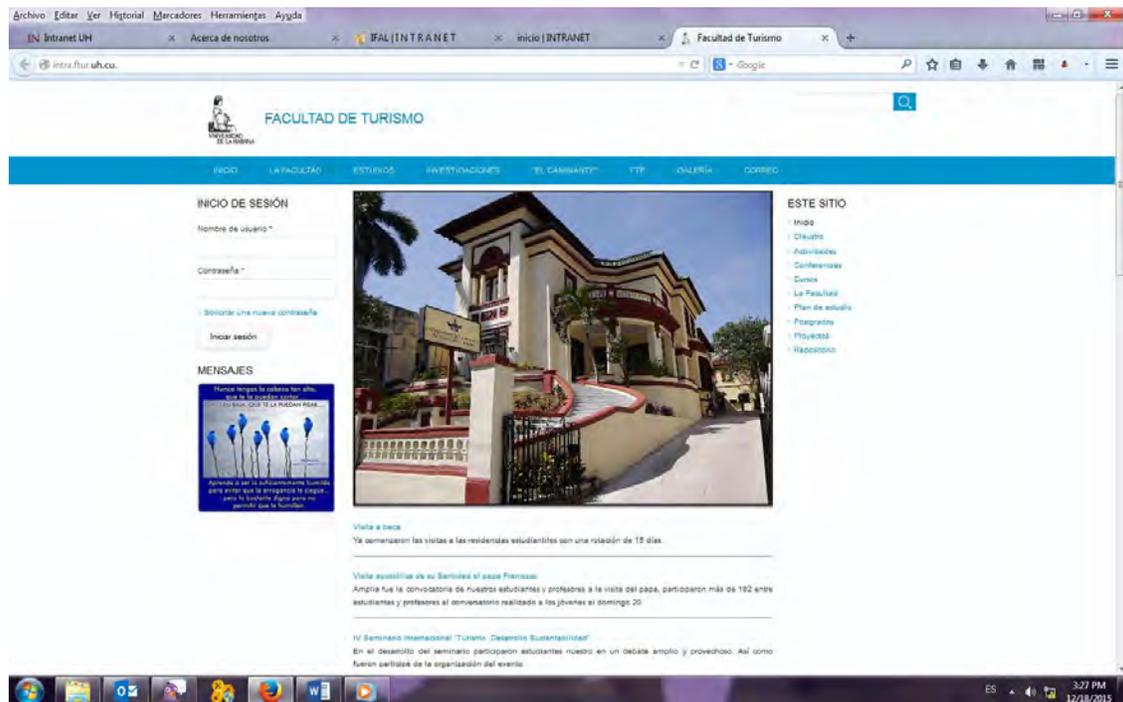


Fig. 5.18. Página de inicio de la intranet de la Facultad de Turismo

5.3.9. Facultad de Filosofía e Historia

La Facultad de Filosofía e Historia, a través de las carreras de Filosofía, Historia y Sociología, prepara a los especialistas capaces para explicar estos procesos en su multiplicidad de factores y condicionantes, prever las dinámicas de desarrollo y realizar los trabajos de diagnóstico e intervención que permitan el perfeccionamiento de nuestro proyecto social.

La Facultad se destaca por su participación en los programas masivos de educación universitaria, y de luchadores sociales, así como por la producción científica de textos para la enseñanza general y superior, programas de investigación de gran alcance y la participación en los cursos

extensionistas "Universidad para todos", que se transmiten por la televisión nacional.

En la Facultad de Filosofía e Historia se estudian las carreras de Licenciatura en Filosofía, Licenciatura en Historia y Licenciatura en Sociología.

La Intranet de la Facultad de Filosofía e Historia (<http://intranet.ffh.uh.cu/>) fue publicada en 2010 (Ver fig.5.19). Es un sitio implementado en Drupal.

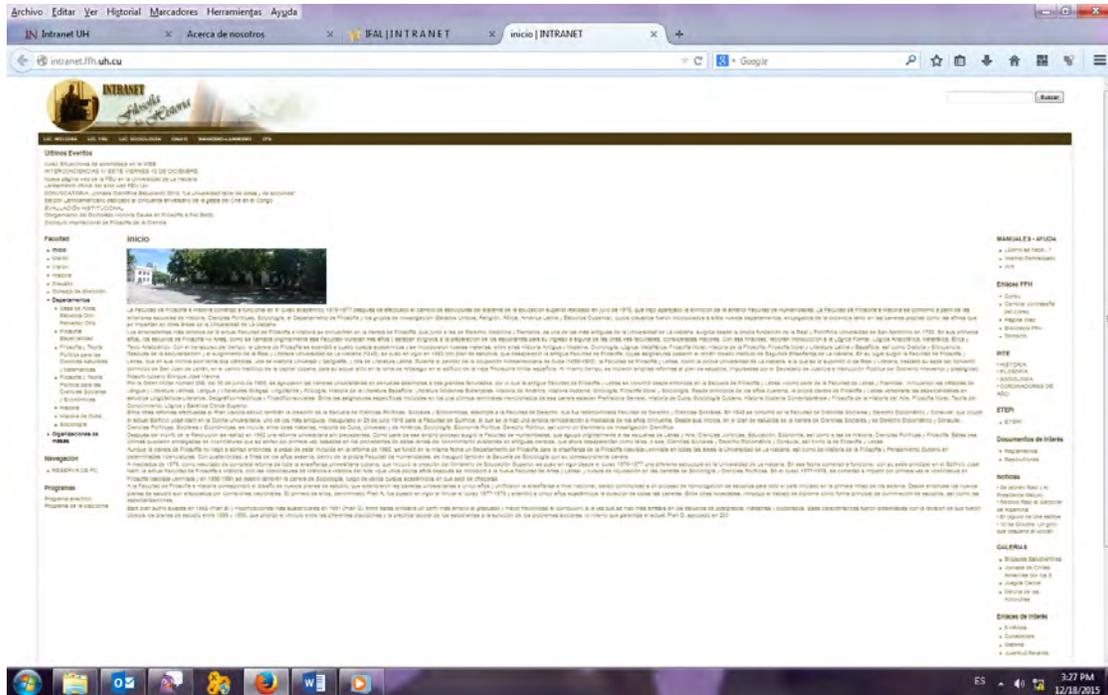


Fig. 5.19. Página de inicio de la intranet de la Facultad de Filosofía e Historia

5.3.10. Facultad de Física

La formación de Licenciados en Física en Cuba comenzó a partir de la Reforma Universitaria de 1962. Estos se preparan para la investigación fundamental o aplicada, así como para los servicios técnicos o la docencia universitaria en el extenso universo temático de la Física. Sus egresados laboran típicamente en el Polo Científico del Oeste, El ICIMAF, el IMRE, la UCI, y otros muchos centros de gran prestigio. La primera edición de la carrera de Ingeniería Física acaba de comenzar en septiembre del 2007. Los ingenieros físicos han de ser capaces de trabajar en la investigación

aplicada, en tareas de control de la calidad, estudios de procesos tecnológicos e introducción de nuevas tecnologías, construcción de equipos y sistemas de medición y desarrollo de software en problemas vinculados con la Física, la tecnología, sus métodos y aplicaciones.

En la Facultad de Física se estudian las carreras de Licenciatura en Física e Ingeniería Física. Su intranet (<http://intranet.fisica.uh.cu/>) fue publicada en 2013 (Ver fig.5.20). Es un sitio implementado en Joomla.

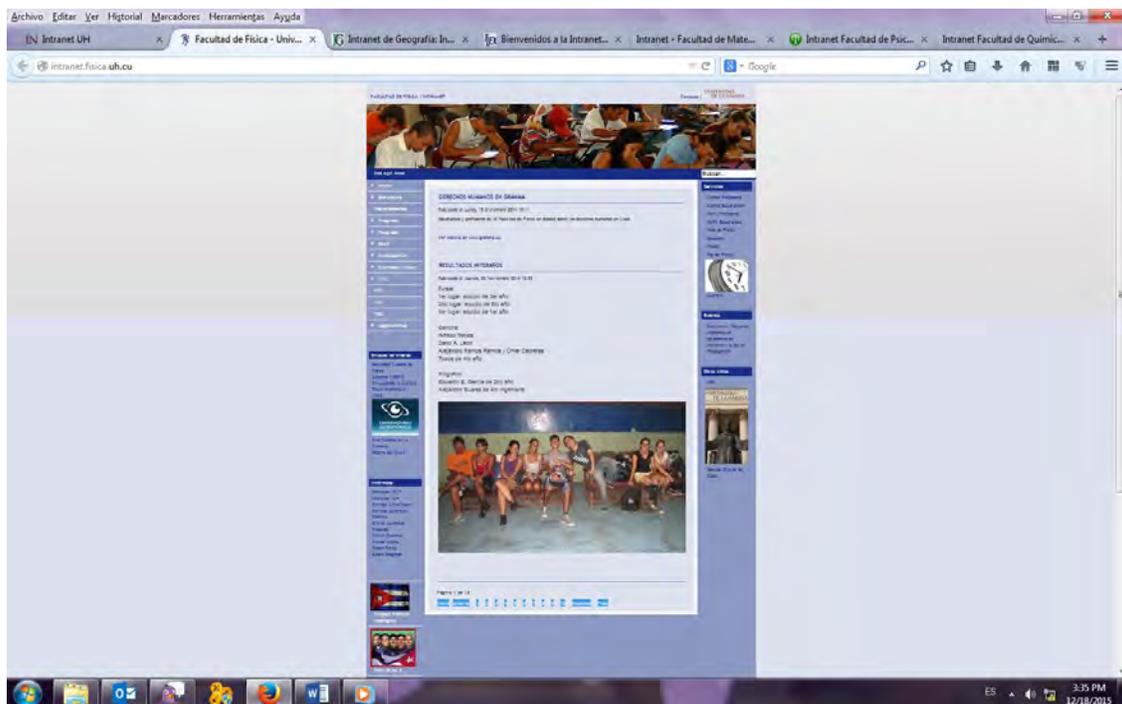


Fig. 5.20. Página de inicio de la intranet de la Facultad de Física

5.3.11. Facultad de Geografía

La carrera de Geografía surge en Cuba como parte de la Reforma universitaria realizada en 1962, momento en que se creó la escuela de geografía de la Universidad de La Habana. Este paso permitió una paulatina intensificación de los estudios en las ciencias geográficas y en particular en la labor científico-investigativa vinculada a las mismas. Desde estos momentos, el trabajo de investigación persiguió 3 objetivos básicos que hoy mantienen plena vigencia:

- Constituir un elemento indisolublemente ligado a la formación del futuro profesional.
- Propiciar el desarrollo y calificación del personal docente para las investigaciones.
- Garantizar el uso del potencial científico que labora y estudia en la facultad para la solución de tareas específicas del desarrollo económico y social del país.

Por ello se ha puesto el énfasis en las investigaciones de carácter aplicado, vinculadas a la solución de problemas sociales, investigaciones que relacionan los trabajos técnicos fundamentales con la práctica, lo que ha sido una vía efectiva y operativa para introducir los resultados de la investigación en la economía, la producción y la organización social.

En la Facultad de Geografía se estudia la carrera Licenciatura en Geografía. Su intranet (<http://intranet.geo.uh.cu/>) fue rediseñada e implementada en 2016(Ver fig.5.21). Es un sitio implementado en Joomla

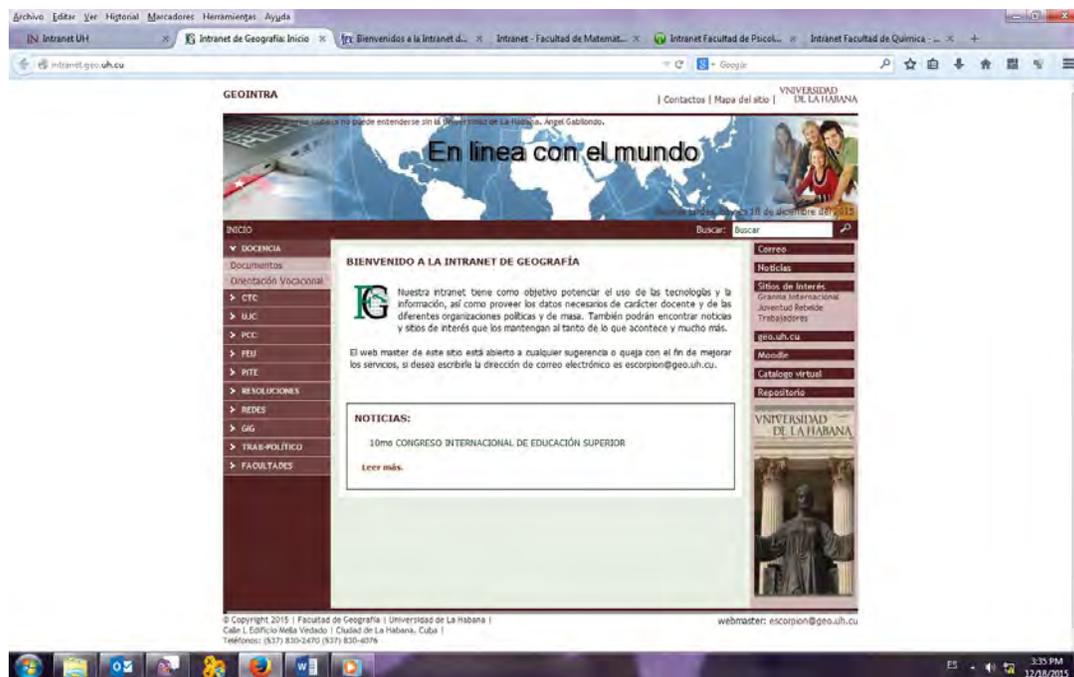


Fig. 5.21. Página de inicio de la intranet de la Facultad de Geografía

5.3.12. Facultad de Lenguas Extranjeras

La Facultad de Lenguas extranjeras capacita a sus estudiantes para la traducción e interpretación, al menos en dos lenguas además de la materna; así como para la docencia de nivel superior, y en general para las actividades vinculadas con las relaciones exteriores, el turismo y otras afines. Prepara, además, a los intérpretes en lengua de señas para el trabajo con sordos e hipoacúsicos.

La Facultad se destaca por la extensión de su doctorado curricular, los resultados de sus estudios semánticos, la captación de divisas y el servicio de superación lingüística a docentes y estudiantes universitarios, así como en los cursos extensionistas de televisión "Universidad para todos".

En la Facultad de Lenguas extranjeras se estudian las carreras de Licenciatura en Lengua Alemana, Licenciatura en Lengua Francesa, Licenciatura en Lengua Inglesa y Licenciatura en Lengua Rusa.

La Intranet de la Facultad de Lenguas extranjeras (<http://intranet.flex.uh.cu/>) fue rediseñada en 2009 (Ver fig.5.22). Es un sitio implementado en Drupal.



Fig. 5.22. Página de inicio de la intranet de la Facultad de Lenguas Extranjeras

5.3.13. Facultad de Matemática y Computación

La Facultad de Matemática y Computación tiene como objetivo formar Licenciados en Matemática y Ciencia de la Computación de perfil amplio, es decir, profesionales capaces de resolver una serie de problemas comunes a varias esferas, aptos para modelar la realidad objetiva que nos rodea y desarrollar softwares de alto nivel profesional.

En la Facultad de Matemática y Computación se estudian las carreras de Licenciatura en Matemática y Licenciatura en Ciencia de la Computación.

La Intranet de la Facultad (<http://intranet.matcom.uh.cu/>) fue rediseñada en 2011 (Ver fig.5.23). Es una página implementada en Bootstrap v3.1.1 (<http://getbootstrap.com>), que es un framework para el desarrollo de proyectos.

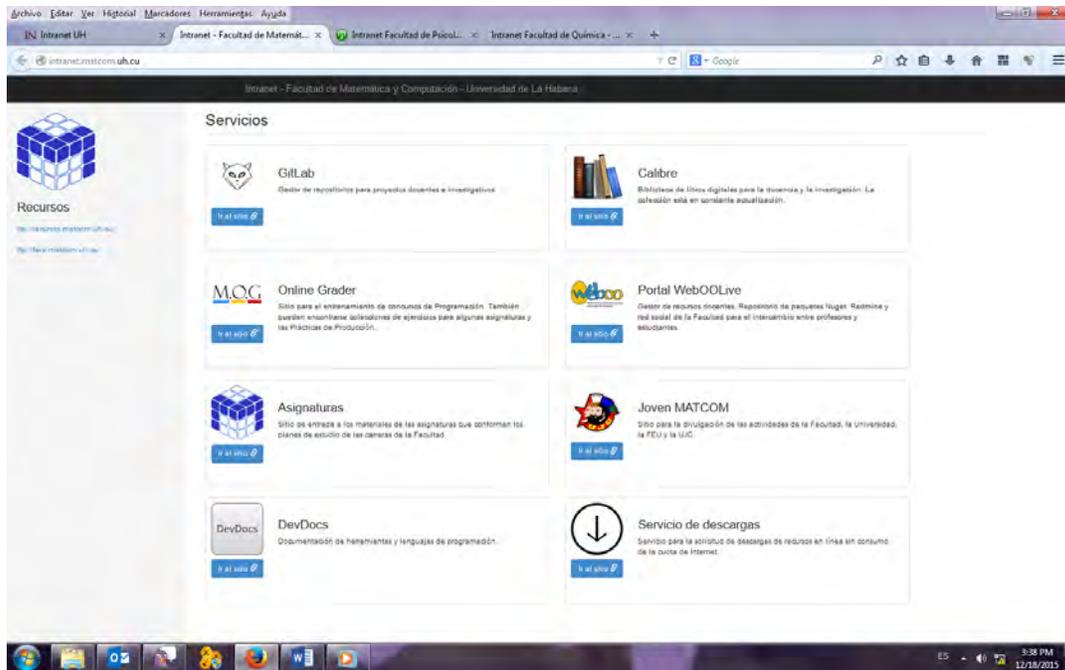


Fig. 5.23. Página de inicio de la intranet de la Facultad de Matemática

Da acceso a un paquete de servicios a los que los diferentes usuarios pueden acceder:

- GitLab: Gestor de repositorios para proyectos docentes e investigativos.
- Calibre: Biblioteca de libros digitales para la docencia y la investigación. La colección está en constante actualización.
- Online Grader: Sitio para el entrenamiento de concursos de Programación. También pueden encontrarse colecciones de ejercicios para algunas asignaturas y las Prácticas de Producción.
- Portal WebOOLive: Gestor de recursos docentes. Repositorio de paquetes Nuget. Redmine y red social de la Facultad para el intercambio entre profesores y estudiantes.
- Asignaturas: Sitio de entrada a los materiales de las asignaturas que conforman los planes de estudio de las carreras de la Facultad.
- Joven MATCOM: Sitio para la divulgación de las actividades de la Facultad, la Universidad, la FEU y la UJC.
- DevDocs: Documentación de herramientas y lenguajes de programación.
- Servicio de descargas: Servicio para la solicitud de descargas de recursos en línea sin consumo de la cuota de Internet

5.3.14. Facultad de Psicología

La Facultad de Psicología prepara a sus especialistas en un amplio espectro de campos tales como familia, mujer, discapacidad, adulto mayor, vinculados con la psicología clínica, educativa, laboral, social, comunitaria.

La Facultad se destaca por su amplia labor asistencial, su presencia en la Universidad municipal y su acción extensionista. En ella se estudia la carrera de Licenciatura en Psicología.

La Intranet de la Facultad (<http://intranet.psico.uh.cu/>) fue rediseñada en 2009 (Ver fig.5.24). Es una página implementada en Drupal

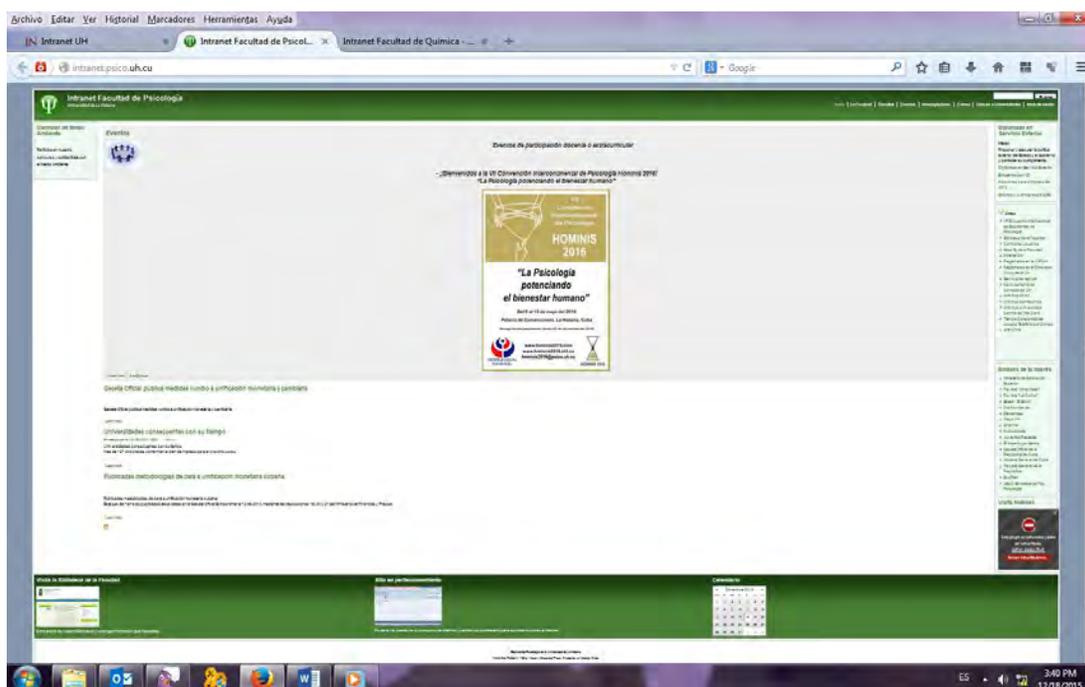


Fig. 5.24. Página de inicio de la intranet de la Facultad de Psicología

5.3.15. Facultad de Química

La Facultad de Química se dedica a la formación integral y continua de especialistas con nivel competitivo; capaces para la realización de investigaciones científicas vinculadas a las direcciones de desarrollo principales de nuestro país, así como para la labor docente y extensionista

en relación con antígenos sintéticos y productos naturales, síntesis orgánica, bioinorgánica, química - física, etc.

La Facultad se destaca por los logros en vacunas y otros productos comercializables, de gran impacto social; por la captación obtenida por estos conceptos, así como por la calidad de su trabajo metodológico, que ha sido de referencia. En ella se estudia la carrera de Licenciatura en Química.

La Intranet de la Facultad (<http://10.6.54.98/fq/index.php/inicio>) fue rediseñada en 2012 (Ver fig.5.25). Es una página implementada en Joomla.

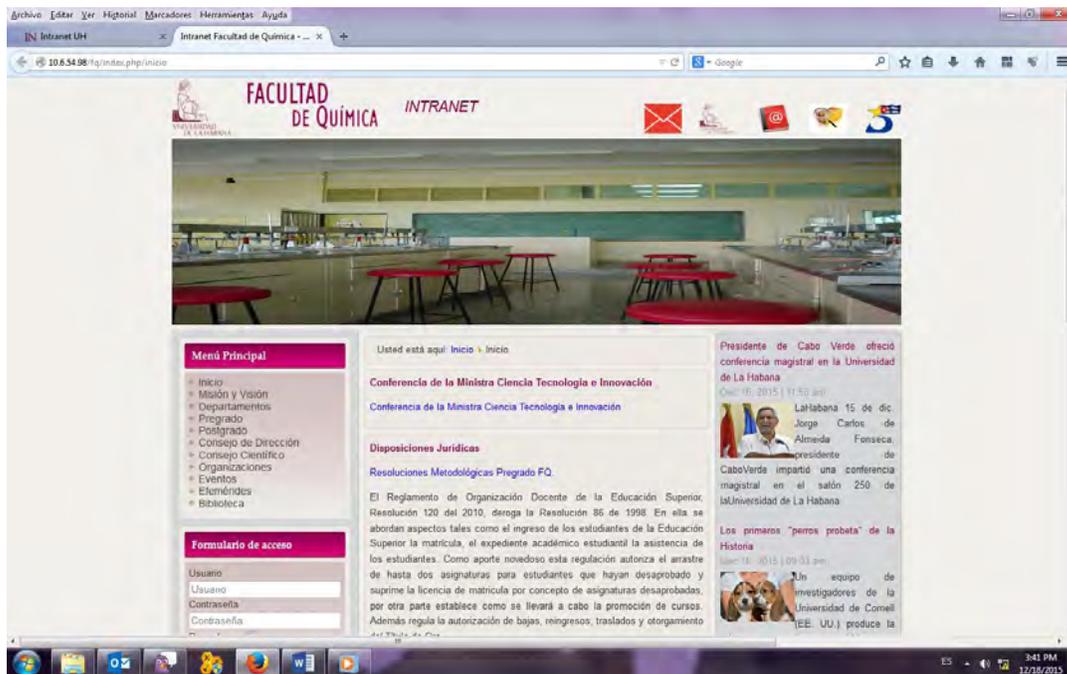


Fig.5.25. Página de inicio de la intranet de la Facultad de Química

5.3.16. Facultad de Español para no Hispanohablantes

La Facultad de Español para No Hispanohablantes (FENHI), surge como programa de la Revolución Cubana, regido por la Universidad de La Habana, con motivo de instituir el idioma español como carrera universitaria para estudiantes de habla no hispana. Es una institución con pocos años de fundada.

Tanto su página Web como la Intranet (<http://intranet.fenhi.uh.cu/>) se encuentran en construcción y para ello utilizan y experimentan con distintos CMS. Las imágenes que se muestran a continuación responden a la construcción de estos productos en Joomla. Actualmente se estudian las potencialidades de Drupal. (Ver Fig.5.26)



Fig. 5.26. Página de inicio de la intranet de FENHI

La Intranet de FENHI, dada la reciente creación de la Facultad, aún no aparece declarada dentro de la Red de Intranets de las Facultades en el portal de la UH.

Aunque la responsabilidad de la actualización de estos espacios recae sobre el Vicedecanato de Investigación y Postgrado de la Facultad y el grupo de informáticos. No obstante, es importante señalar que la actualización de los productos que se generan en FENHI no es encargo exclusivo de los informáticos, pues cada estructura debe montar los contenidos que considere (dos profesores por departamento, dos especialistas por biblioteca, y estudiantes. Los informáticos se ocupan de elementos generales de la Intranet y la Web).

5.3.17. Facultad de Educación a Distancia

La Facultad fue creada con el objetivo de ampliar las posibilidades de estudio de nivel superior (Universitario) de la población cubana, surge esta modalidad de educación en Cuba en el curso académico 1979-1980, coincidiendo esta fecha con la década de auge de la Educación a Distancia a nivel mundial.

Esta Facultad amplía las oportunidades de acceso a la universidad y contribuye a la formación de una cultura integral de nuestra población y a elevar la equidad y justicia social en nuestra sociedad. Se ha caracterizado por el rigor organizativo y el ofrecimiento de facilidades a sus estudiantes.

La Facultad de Educación a Distancia de la Universidad de La Habana ha desarrollado el modelo de educación a distancia durante más de 25 años, como Centro Rector de la actividad.

En la Facultad de Educación a Distancia se estudian las carreras de Licenciatura en Derecho, Licenciatura en Historia, Licenciatura en Contabilidad y Finanzas, Licenciatura en Economía, Licenciatura en Bibliotecología y Ciencia de la Información, Licenciatura en Estudios Socioculturales y Licenciatura en Turismo.

La Intranet de la Facultad (<http://www.fed.uh.cu/index>) fue diseñada en 2007(Ver fig.5.27). Es una página implementada en HTML con Dreamweaver.



Fig. 5.27. Página de inicio de la intranet de la Facultad de Educación a Distancia

5.3.18. CEPES

El Centro de Estudios para el Perfeccionamiento de la Educación Superior de la Universidad de La Habana (CEPES) se creó en el año 1982 a partir de la unificación de diversas áreas de estudio con tradición en la esfera universitaria. Orienta su actividad sistemáticamente hacia la elevación de los niveles de excelencia académica, pertinencia social y cooperación nacional e internacional. Constituye el primer centro con perfil multidisciplinario de investigación y posgrado que sobre este campo de estudio temático se creó en Cuba integrando diferentes perspectivas, entre ellas, las pedagógicas, sociológicas, de gestión y las nuevas tecnologías de la información y las comunicaciones (CEPES, 2012).

La Intranet del CEPES (<http://intra.cepes.uh.cu/>) fue diseñada en 2015 (Ver fig.5.28). Es una página implementada en Drupal



Fig. 5.28. Página de inicio de la intranet del CEPES

5.3.19. FLACSO

La FLACSO Facultad latinoamericana de Ciencias Sociales (FLACSO) es un organismo internacional, intergubernamental, de carácter regional y autónomo, creado desde 1957 y constituido por 14 países latinoamericanos y caribeños, para promover la enseñanza y la investigación en el campo de las Ciencias Sociales. Constituye, hasta hoy día, el mayor y más importante programa regional de formación en Ciencias Sociales. Sus unidades actúan en diez países del área, que realizan labores de investigación y docencia post graduada en diversas temáticas (FLACSO, 2014).

La Intranet del FLACSO (<http://intranet.flacso.uh.cu/>) fue diseñada en 2008(Ver fig.5.29). Es una página implementada en Drupal.



Fig. 5.29. Página de inicio de la intranet del FLACSO

5.3.20. CEAP

El Centro de Estudios de Administración Pública tuvo sus inicios en la Cátedra de Administración Pública que se subordinaba a la Vicerrectoría de Relaciones Internacionales y Postgrado desde el año 1997. El Centro como tal se crea el 21 de abril de 2002 por la Resolución Rectoral 70/02, del Ministro de Educación Superior como un Centro científico-investigativo y docente de la Universidad de la Habana con el objeto de desarrollar acciones de superación profesional y formación académica de postgrado, investigaciones, actividades de asesoría, proyectos y convenios de colaboración entre otras (CEAP, 2015).

La Intranet del CEAP (<http://www.ceap.uh.cu/>) fue diseñada en 2013 (Ver fig.5.30). Es una página implementada en Joomla



Fig. 5.30. Página de inicio de la intranet del CEAP

5.3.21. CEEC

El Centro de Estudios de la Economía Cubana (CEEC) fue fundado en 1989, cuya misión es contribuir al estudio y desarrollo de la Economía Cubana a través de la docencia, la investigación, la capacitación y la prestación de servicios de consultoría en los ámbitos de la economía y la gerencia (CEEC, 2016).

Mantiene relaciones de colaboración, capacitación y asesoría con varios Organismos de la Administración Central del Estado, la Cámara de Comercio de la República de Cuba y el Consejo de Administración Provincial CAP de Ciudad de la Habana (CEEC, 2016).

La Intranet del CEEC (<http://intranet.ceec.uh.cu/>) fue diseñada en 2007(Ver fig.5.31). Es una página implementada en Drupal.



Fig. 5.31. Página de inicio de la intranet del CEEC

5.3.22. BIOMAT

El Centro de Biomateriales (BIOMAT), fue creado el 23 de octubre de 1991 como centro de investigación cuya misión es la producción de conocimientos científicos y tecnológicos en el campo de los materiales de uso médico, el desarrollo y la producción de biomateriales y productos químicos para el diagnóstico y las investigaciones biomédicas y su introducción en la asistencia de salud cubana. Su actividad investigativa incluye desde la búsqueda de soluciones a problemas teóricos de las Ciencias de los Biomateriales, hasta la obtención de productos comercializables.

BIOMAT lidera en el país los estudios, la innovación tecnológica y la gestión de la calidad de los biomateriales poliméricos, cerámicos y compuestos.

La Intranet del BIOMAT (<http://intranet2.biomat.uh.cu/>) fue diseñada en 2014 (Ver fig.5.32). Es una página implementada en Drupal.



Fig. 5.32. Página de inicio de la intranet de BIOMAT

5.3.23. CIEI

Fundado en 1970, el Centro de Investigaciones de Economía Internacional de la Universidad de La Habana deviene una institución académica que tiene como objetivo fundamental contribuir, a través de la investigación y la docencia, a una mejor comprensión de los procesos de la economía internacional y de sus efectos sobre los países subdesarrollados.

La misión institucional del Centro se resume en obtener y divulgar los resultados de investigación de especialistas de alto nivel, cuyo colectivo de autores contribuye al perfeccionamiento de una base nacional de conocimientos teóricos en el área de las Relaciones Económicas Internacionales.

La Intranet del CIEI (<http://intranet.ciei.uh.cu/>) fue diseñada en 2007 (Ver fig.5.33). Es una página implementada en Drupal.

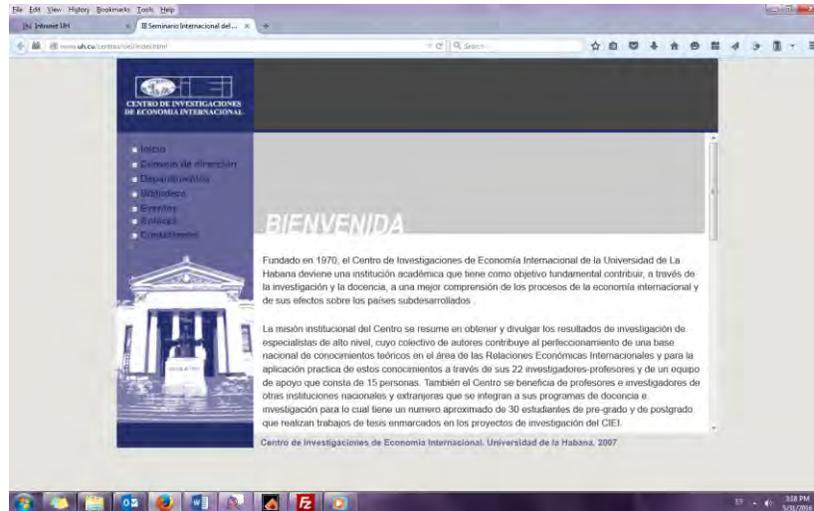


Fig. 5.33. Página de inicio de la intranet del CIEI

5.3.24. CEDEM

El Centro de Estudios Demográficos de la Universidad de La Habana (CEDEM) fue creado en febrero de 1972 como institución dedicada a las investigaciones en el campo del análisis demográfico y de los estudios de población y desarrollo sostenible (CEDEM, 2012).

Sus resultados, de elevado impacto social, contribuyen al progreso de la sociedad cubana. Su actividad científica aporta, por ejemplo, a la solución de problemas vinculados a las migraciones internas y su labor es decisiva en la organización y ejecución de los censos nacionales y en otras investigaciones que sobre población se realizan en el país (CEDEM, 2012).

La Intranet del CEDEM (<http://intranet.cedem.uh.cu/>) fue diseñada en 2013(Ver fig.5.34). Es una página implementada en Drupal.

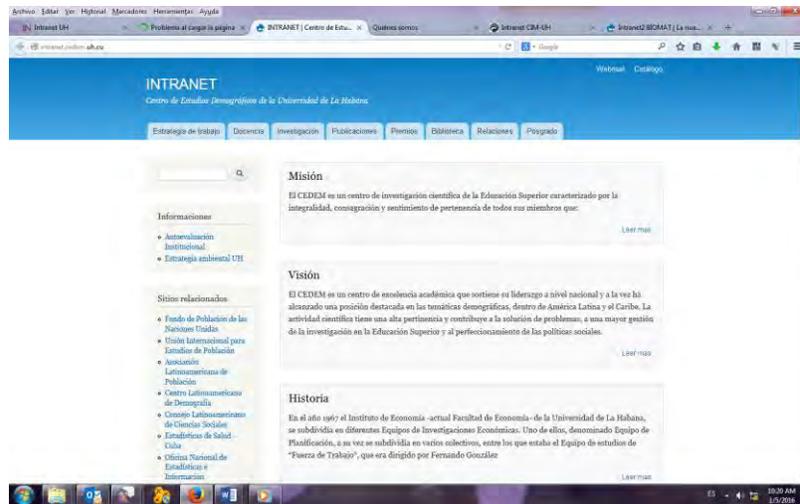


Fig. 5.34. Página de inicio de la intranet del CEDEM

5.3.25. CIM

El Centro de Investigaciones Marinas de la Universidad de La Habana (CIM-UH) tiene la misión de contribuir a la conservación del ambiente y al desarrollo sostenible a través de la integración de la investigación científica y la formación continua de los profesionales en Biología Marina, Acuicultura y Manejo Costero, con enfoque multidisciplinario y de excelencia (CIM, 2015).

Para cumplir esa Misión, el CIM-UH cuenta con más de 35 años de experiencia en la formación de especialistas en Biología Marina y Acuicultura en pregrado y postgrado a través de cursos y entrenamientos, no sólo para estudiantes cubanos sino también de otros países de América, África, Europa y Asia. En este empeño cuenta con la colaboración de especialistas de alto nivel procedentes de otros centros de investigación como el Instituto de Oceanología del CITMA y el Centro de Investigaciones (CIM, 2015).

La Intranet del CIM (<http://intranet.cim.uh.cu/>) fue diseñada en 2007 (Ver fig.5.35). Es una página implementada en Drupal.



Fig. 5.35. Página de inicio de la intranet del CIM

5.3.26. IFAL

La Facultad de Farmacia y Alimentos asegura la formación de los especialistas capacitados para la tecnología y control de medicamentos, química farmacéutica, farmacología, alimentación y nutrición, procesamiento y evaluación de alimentos, entre otras materias, y en estrecha coordinación con los Centros del Polo Científico, y otros organismos como el MINSAP.

La Facultad se destaca por los programas de prevención y lucha contra las adicciones y el SIDA, así como por otros proyectos comunitarios relacionados con el uso y control de medicamentos, y con la nutrición.

En la Facultad de Farmacia y Alimentos se estudian las carreras de Licenciatura en Ciencias Farmacéuticas y Licenciatura en Ciencias Alimentarias (IFAL, 2012)

La Intranet del IFAL (<http://intranet.ifal.uh.cu/>) fue diseñada en 2007 (Ver fig.5.36). Es una página implementada en Drupal

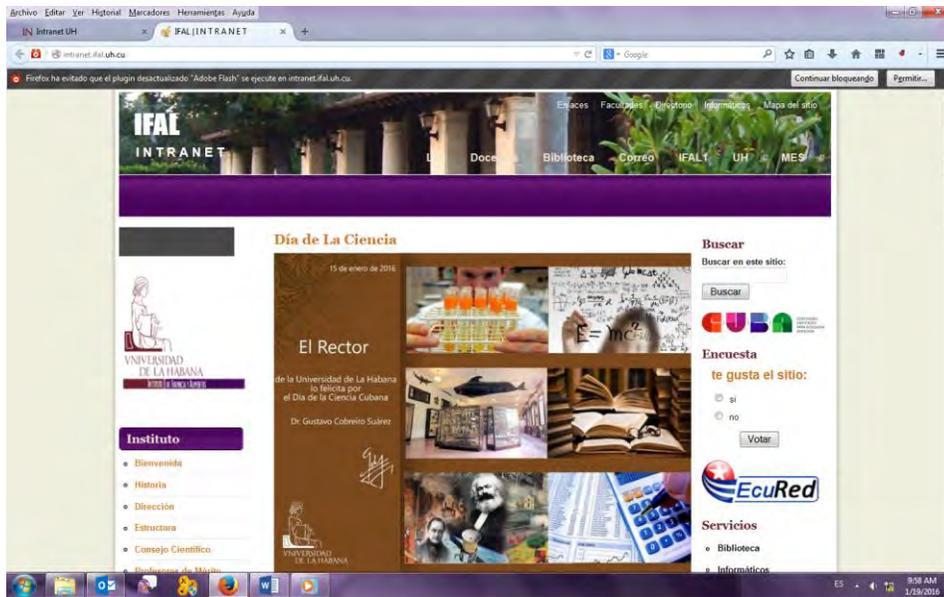


Fig.5.36. Página de inicio de la intranet IFAL

5.3.27. Análisis del uso de los CMS para la construcción de Intranets

Se contactó con los administradores de red de cada facultad, porque son las personas que desempeñan la función de implementación y mantenimiento de las intranets del área correspondiente, al asumir la responsabilidad de las políticas informáticas del centro, se convierte en la indicada para desempeñar el rol de administrador de la Intranet y/o el sitio Web y, con ello, asume la responsabilidad de actualización de estos productos.

Del total de estructuras de la UH que posee una Intranet (26 estructuras), el 80% (21) están implementadas en algún tipo de CMS. (Ver Fig. 5.37)

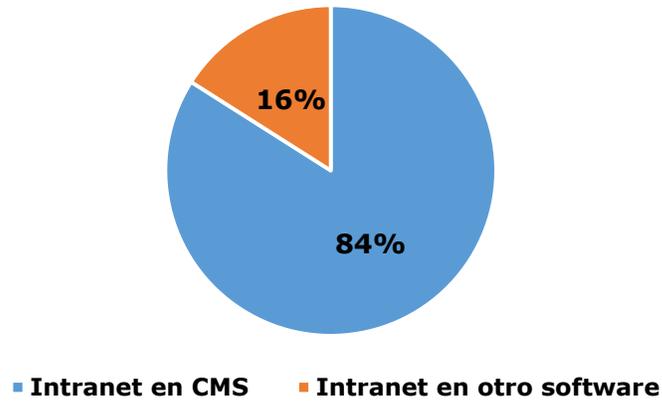


Fig. 5.37. Intranet de la UH y proporción de Software utilizados, enero/ 2016 (Elaboración Propia)

Se identificaron, en enero de 2016, un total de 3 CMS, todos de tipo software libre. (Ver Fig.5.38)

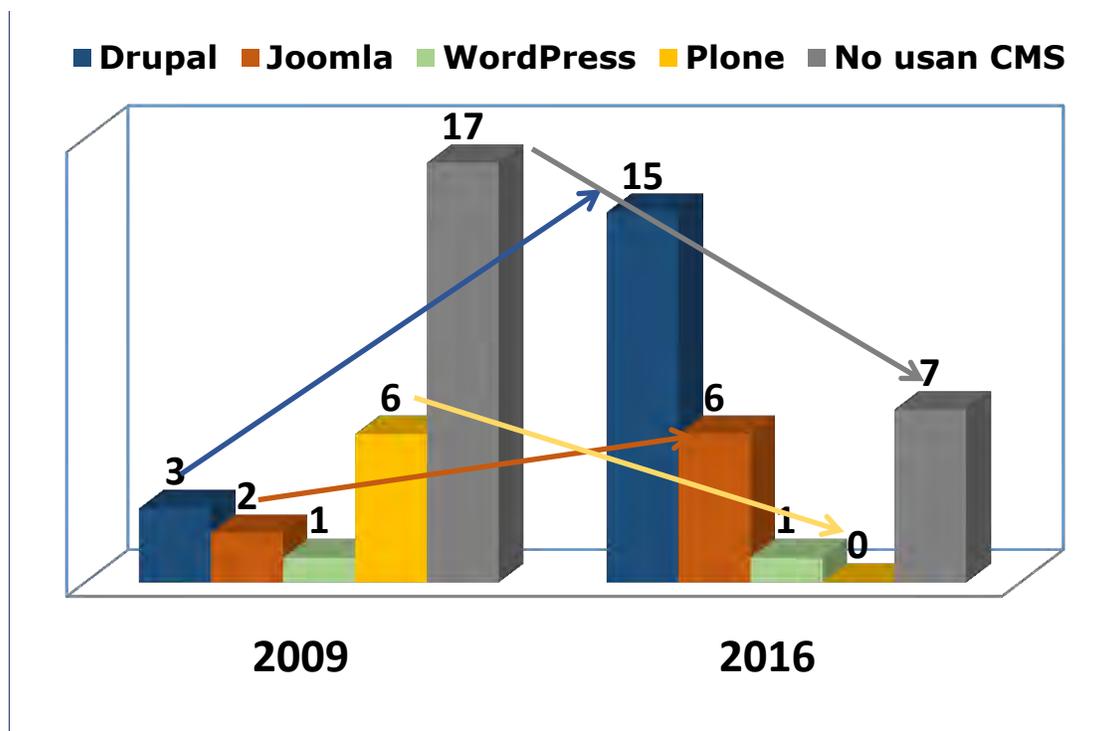


Fig. 5.38. CMS identificados en las Intranets de la UH

Como puede apreciarse en la gráfica, en el período de 5 años, ha habido un incremento considerable en el uso de CMS en la UH, especialmente de Drupal y Joomla. Dicha tendencia está en correspondencia con los avances en cuanto a las potencialidades de usabilidad, amigabilidad de software que

han tenido las nuevas versiones de Drupal y Joomla, así como el crecimiento de sus comunidades de desarrollo.

Es válido señalar que en el caso de las intranets que hoy no se encuentran implementadas sobre un CMS se corresponden con las áreas que no cuentan con personal informático o personal designado al mantenimiento y desarrollo de la intranet.

Las intranets se caracterizan por el uso de un gran número de documentos en versión digital, de calidad variable y temática multidisciplinar.

La naturaleza de los documentos no siempre es textual. Hay gran cantidad de imágenes, y en menor proporción videos y documentos de audio. Dadas las características de los sistemas de recuperación de información implementados, el sistema tiene fallos relevantes, especialmente con los documentos no textuales, pues la indización se basa en la indización de palabras que forman parte del texto del documento. En consecuencia se impone la necesidad de un mecanismo alternativo de indización que en la UH de hoy no existe.

El grado de interoperabilidad entre las intranet es escaso y en consecuencia el nivel de reutilización de la información.

Por otra parte con el aumento de sitios construidos con CMS, se nota un cambio no solo en el uso de la herramienta, también en el papel de los usuarios, quienes han dejado de ser y comportarse como meros espectadores o recuperadores de la información, actualmente han pasado a tomar un papel más activo en la creación, publicación y compartición de recursos.

5.4. Sistemas de gestión de información institucional automatizados de la UH

La Universidad cuenta con varios sistemas de gestión, accesibles desde la intranet, creados a partir del interés de la dirección universitaria de agilizar y transparentar los procesos de gestión en las diferentes áreas de acción institucional. En su mayoría fueron diseñados e implementados desde el

grupo de trabajo IRIS, con la colaboración de las áreas interesadas y la Dirección de Informatización de la UH. Esa es la razón por lo que en su mayoría están implementados en Drupal.

Actualmente se utiliza la versión de Drupal 6.0 y se está estudiando las posibilidades para migrar a Drupal 8 por las posibilidades semánticas que ofrece (Ruiz Jhones, 2016).

El sistema de interacción entre los diferentes sistemas (Ver fig. 5.39) es bastante lineal y simple a nivel comunicacional (Aguilera, 2016).

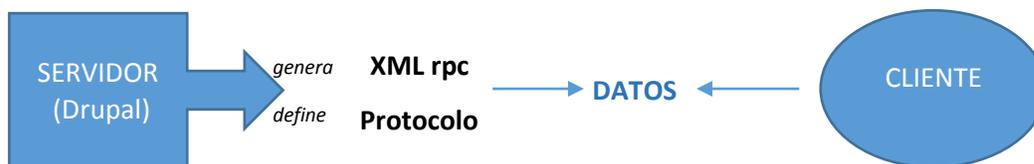


Fig. 5.39. Esquema de comunicación entre los sistemas de gestión en la UH

Los sistemas se abordarán a continuación, refiriendo las principales características de cada uno: objetivo, arquitectura, software de implementación y características de los datos que maneja.

5.4.1. Autenticador Único UH

Como parte de la política de desarrollo de la UH, alrededor del 2010, comenzaron a diseñarse diferentes sistemas que facilitarían la gestión institucional de la universidad. Junto con la Intranet se diseñó e implementó (2011) un sistema cuyo objetivo es la identificación de los usuarios: Autenticador Único UH (ver fig. 5.40) o Directorio Único como también se le conoce (Single SingOn).

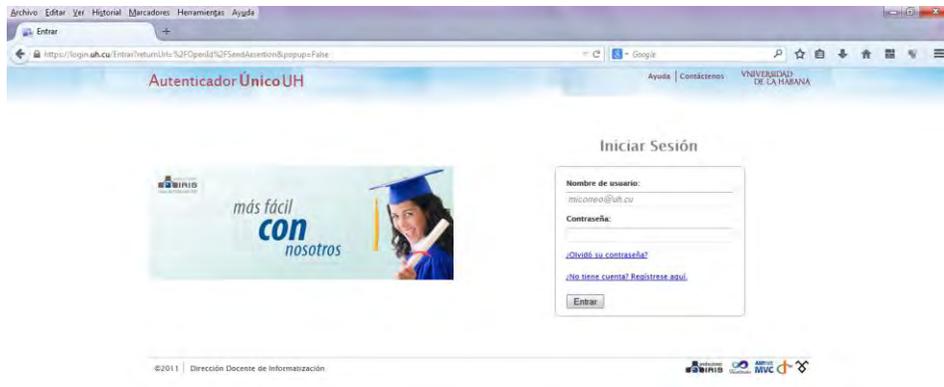


Fig. 5. 40. Visualización de la página de inicio de acceso al sistema Autenticador Único UH o Directorio Único UH

Construido bajo la plataforma ASP.net MVC. Es un Open ID que exporta los datos de los usuarios al resto de los sistemas cuando lo requieren por medio del Webservice SOAP (ver Fig. 5.41) como estrategia para compartir datos, lo que posibilita que el usuario no tiene que autenticarse de diferente manera para cada uno, con el registro en el Directorio único es suficiente.

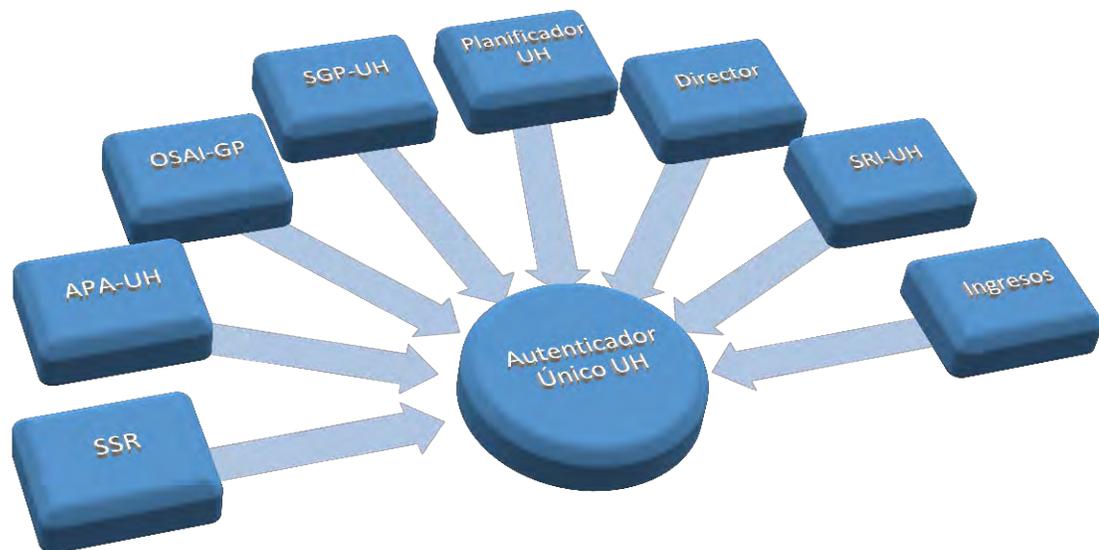


Fig. 5. 41. Comunicación de los sistemas UH con el directorio Único (elaboración propia)

De la misma manera se nutre de los datos aportados por las bases de datos de Recursos Humanos de la UH (construida en MySQL Server, contiene el registro de los trabajadores) y del sistema SIGENU (implementado en Java,

contienen los datos relativos a las matrículas de estudiantes de la universidad) (Ver Fig. 5.42 y 5.43).

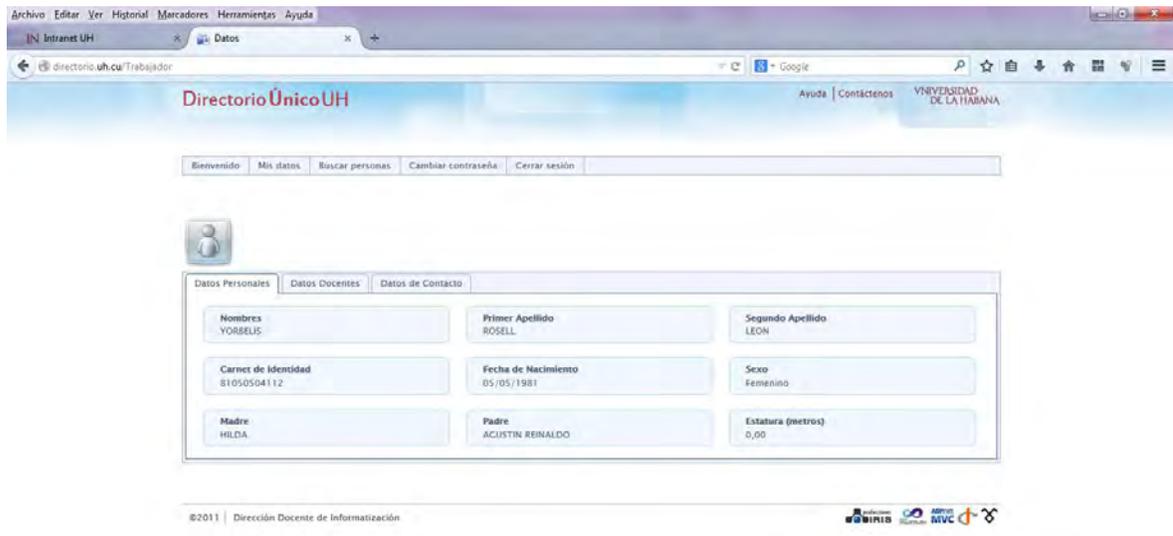


Fig. 5. 42. Ejemplo 1 de registro de un usuario en el Directorio Único: Datos personales

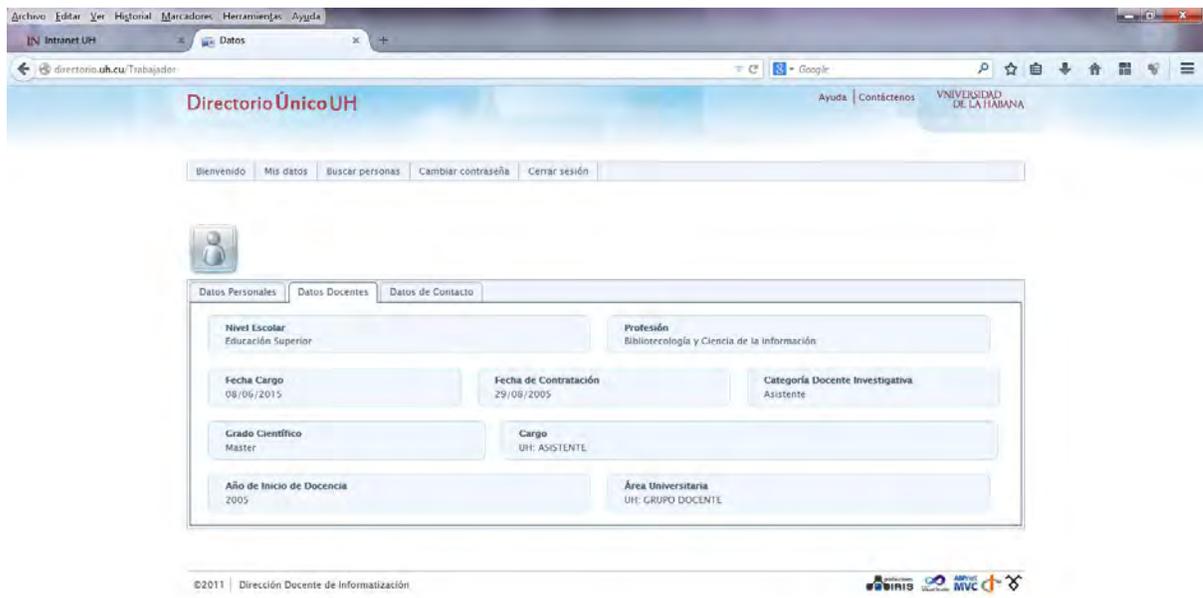


Fig. 5. 43. Ejemplo 2 de registro de un usuario en el Directorio Único: Datos docentes
Por las mismas razones el sistema permite realizar búsquedas booleanas simples sobre estudiantes y trabajados de la UH. (Ver Fig. 5.44 y 5.45)

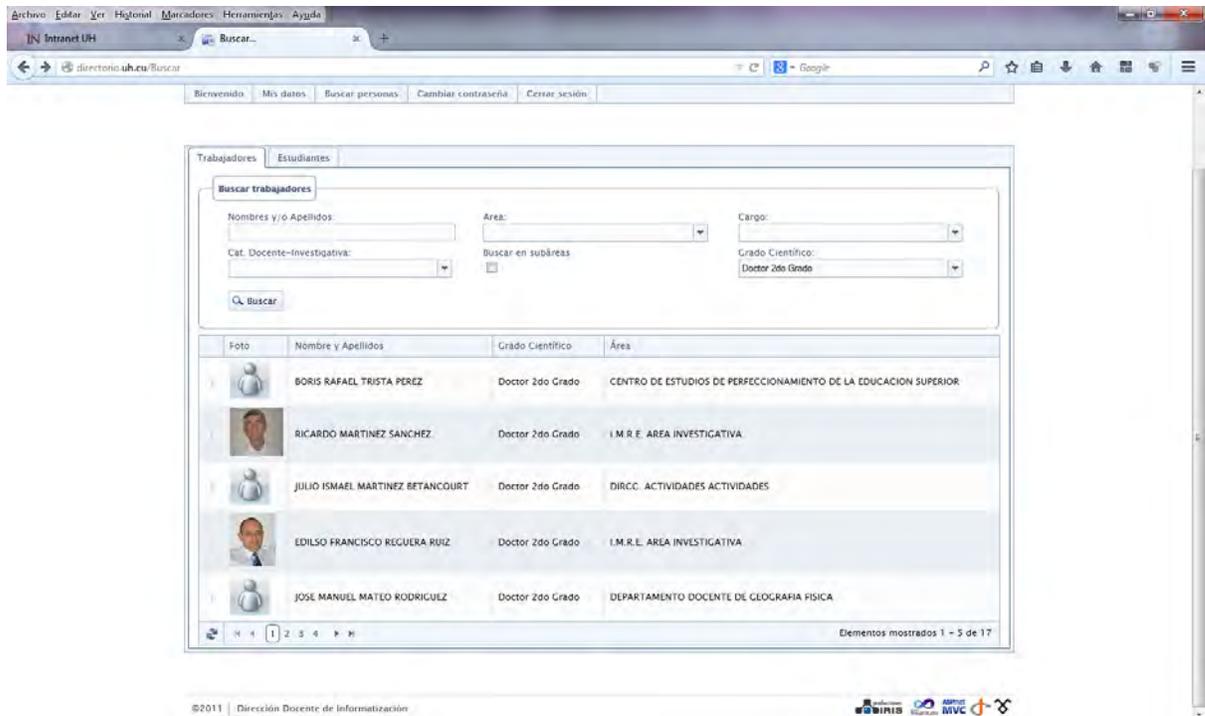


Fig. 5. 44. Directorio Único: búsqueda de trabajadores

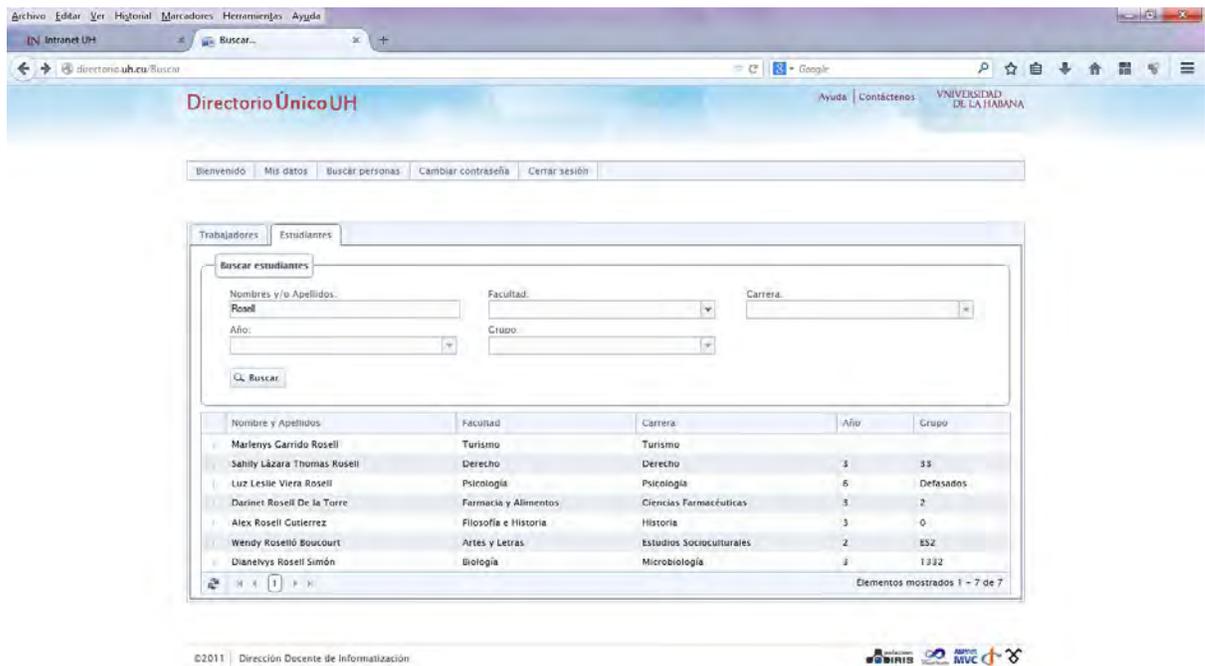


Fig. 5. 45. Directorio Único: búsqueda de estudiantes

5.4.2. Consejo universitario

Objetivo: El sitio web del Consejo Universitario funciona como gestor de eventos y documentos de interés para sus miembros e invitados. Para acceder a esos contenidos debe tener un usuario. (Ver Fig. 5.46)

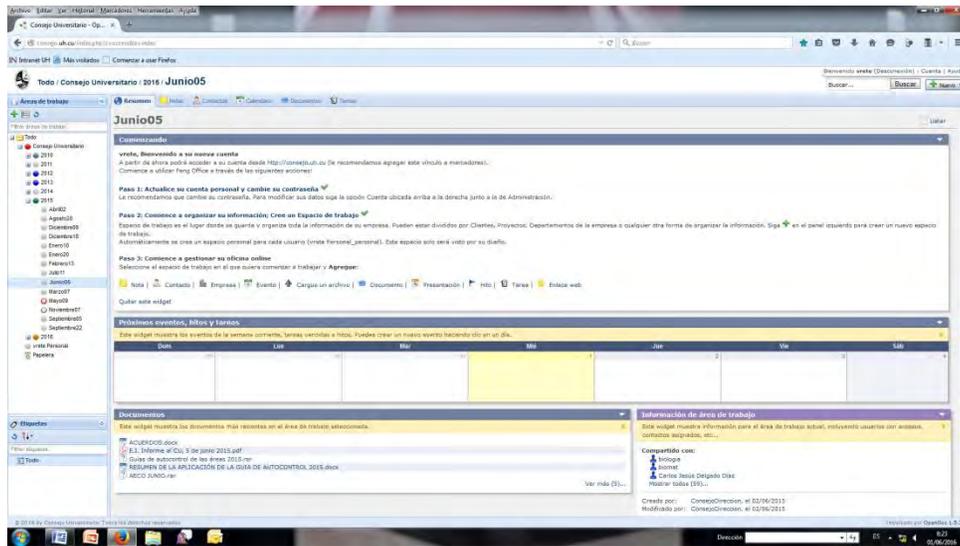


Fig. 5.46. Página de inicio del sitio del Consejo Universitarios de la UH

Software: Drupal

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP.

5.4.3. Departamento jurídico

Objetivo: Departamento Jurídico de la Universidad de La Habana, repositorio de disposiciones jurídicas que brinda información sobre los procedimientos, políticas, regulaciones y funciones que se llevan a cabo en nuestro Centro. Con el objetivo de difundir las regulaciones que rigen la actividad universitaria en sentido general, usted puede acceder a las mismas y descargarlas para su consulta y lectura. Se han organizado por categorías para un mayor entendimiento, no obstante nuestro equipo de asesores se encuentra en disposición de perfeccionar esta herramienta escuchando sugerencias que permitan un constante mejoramiento y actualización del sitio (Ver Fig. 5.47).



Fig. 5.47. Página de inicio del sitio Departamento Jurídico

Software: Drupal

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP.

5.4.4. Sistema de Relaciones Internacionales (SRI-UH)

Objetivo: El Sistema de Relaciones Internacionales (SRI-UH) es el encargado de la gestión de la información referente a la salida y entrada de cubanos y extranjeros desde o hacia la universidad. (Ver Fig 5.48)



Fig.5.48. Página de inicio del sitio SRI-UH

Software: Symfony (en la versión preparada para Drupal)

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP.

5.4.5. Sistema de Asignación de Responsables (Director)

Objetivo: Director lleva a cabo el proceso de asignación de responsables en diferentes sitios. De manera rápida y sencilla, el jefe selecciona a los subordinados y automáticamente estos se adicionan como usuarios en el sistema especificado, luego las personas asignadas pueden visitar el sitio y comenzar a trabajar en él. (Ver Fig. 5.49)



Fig. 5.49. Página de inicio de Director

Software: Drupal

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP. El sistema utiliza RSS

5.4.6. Sistema de Gestión para la Disponibilidad Financiera de la Universidad de La Habana (Disponibilidad Financiera)

Objetivo: Gestionar de manera transparente y de forma digital los datos sobre los recursos financieros a disposición de las diferentes áreas de la UH (Ver Fig. 5.50)



Fig. 5.50. Página de inicio del sitio de gestión de Disponibilidad Financiera

Software: Drupal

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP. El sistema utiliza RSS

5.4.7. Sistema de gestión económica-financiera (EkoWeb)

Espacio de la Dirección de Economía de la Universidad de La Habana que presta servicios económicos, contables y financieros veraces, oportunos y relevantes como apoyo de la toma de decisiones en esta esfera a la comunidad universitaria y a otras instituciones relacionadas con esta actividad. (Ver Fig. 5.51 y 5.52)

Construido en Joomla! y valido XHTML y CSS.



Fig. 5.51. Página de inicio del sitio Eko-Web



Fig. 5.52. Página de inicio para la gestión on-line

Software: Joomla

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP. El sistema utiliza RSS

5.4.8. Sistema de Gestión de Ingresos de la Dirección de Economía (Ingresos)

Objetivo: Gestionar de manera transparente los datos sobre los ingresos y préstamos realizados a las diferentes áreas de la UH (Ver Fig.5.53)

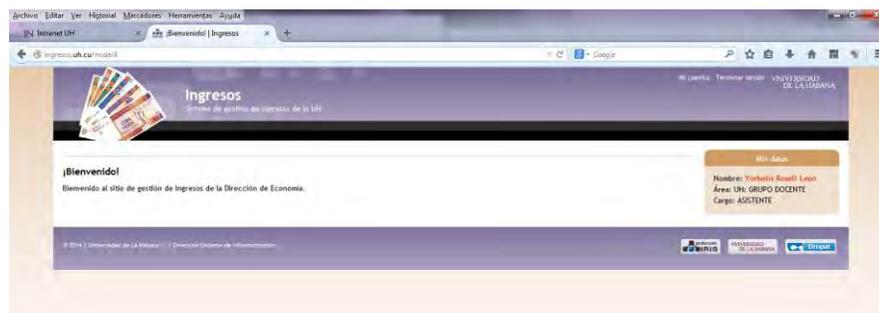


Fig.5.53. Página de inicio de Ingresos

Software: Drupal

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP. El sistema utiliza RSS

5.4.9. Planificador UH

Objetivo: Herramienta para la confección y divulgación del Plan de Trabajo del Rector de la Universidad de La Habana. Para acceder a este servicio debe poseer un usuario (ver Fig 5.54 y Fig. 5.55)

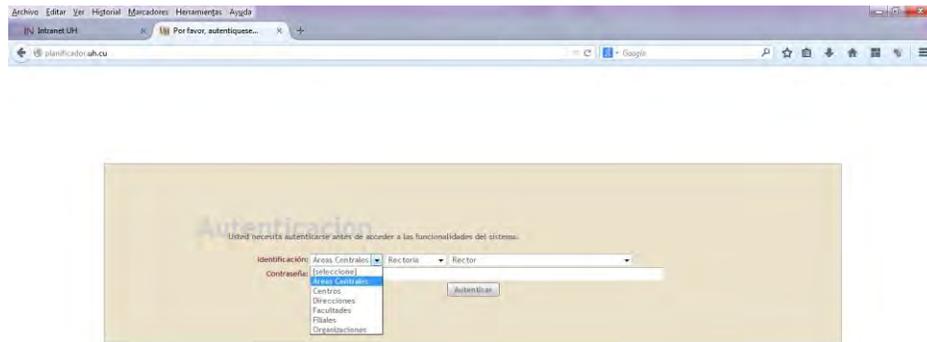


Fig. 5.54. Página de inicio para entrar al Planificador



Fig. 5.55. Página de trabajo

Software: Drupal

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP. El sistema utiliza RSS

5.4.10. Servicios Académicos Internacionales (OSAI-GP)

Objetivo: Sitio de la Oficina de Servicios Académicos Internacionales, "OSAI-GP", el cual permite la gestión de pagos de los cursos ofrecidos a estudiantes foráneos. (Ver Fig.5.56)



Fig. 5.86. Página de inicio de OSAI-GP

Software: Drupal

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP. El sistema utiliza RSS

5.4.11. Sistema de Gestión de la Información de las investigaciones de la Universidad de La Habana (SGI-UH)

Objetivo: El Sistema de Gestión de la Información de las Investigaciones de la Universidad de La Habana (SGI-UH), fue desarrollado por la Dirección de Informatización y la Vicerrectoría de Investigaciones. Tiene como objetivo gestionar la información básica de los proyectos de investigación y sus resultados. Con la finalidad de que todos los trabajadores universitarios conozcan los proyectos y los resultados de la investigación universitaria (artículos, libros, patentes, etc.).(Ver 5.57)



Fig. 5.57. Página de inicio del SGI-UH

Software: Drupal

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP. El sistema utiliza RSS

5.4.12. Sistema de Gestión de la Nueva Universidad (SIGENU)

Objetivo: El Sistema de Gestión de la Nueva Universidad (SIGENU) es un sistema que se ha desarrollado con el fin de ser una herramienta que permita la gestión de toda la información académica vinculada con la educación superior en Cuba. En correspondencia con su carácter nacional y la gran diversidad de sistemas de enseñanza superior con que cuenta la universidad cubana, este sistema ha sido concebido de manera tal que sea capaz de brindar gran seguridad e integridad de la información, y a la vez, ser tan flexible que permita ser adaptado a todos los centros de educación superior del país con sus diversas particularidades y distintas maneras de realizar determinados procedimientos. (Ver Fig.5.58)



Fig. 5.58. Página de inicio del SIGENU

Software: Java. (Diseñado en la CUJAE e implementado en la UH por sus especialistas)

Fuentes y datos: datos primarios. Se introducen directamente con el proceso de matrícula. El sistema trabaja con datos en formato REST y XML rpc.

5.4.13. Sistema de Gestión de Postgrado (SGP-UH)

Objetivo: Sistema de Gestión de Postgrados (SGP) de la Universidad de La Habana, desarrollado por la Dirección de Informatización y la Vicerrectoría de Investigaciones y Postgrado. (Ver Fig. 5.59)

En este sistema los usuarios encontrarán información sobre los requisitos que deben cumplir para integrarse a cualquiera de nuestras formas organizativas de posgrado, también se muestran las opciones de cursos, maestrías, doctorados y pueden solicitar su matrícula.

Los profesores y directivos pueden inscribir cualquiera de las formas de superación posgraduada y gestionarlas posteriormente.



Fig. 5.59. Página de inicio al SGO-UH

Software: Drupal

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP. El sistema utiliza RSS

5.4.14. Sistema de Solicitud de Recursos (SSR)

Objetivo: Sistema de Solicitud de Recursos (SSR) de la Universidad de La Habana, desarrollado por la Dirección de Informatización y la Vice Rectoría Docente. Tiene como objetivo gestionar la información básica de los locales, así como las reservaciones de los mismos. (Ver Fig. 5.60 y Fig. 5.61)



Fig. 5.60. Página de inicio de SSR

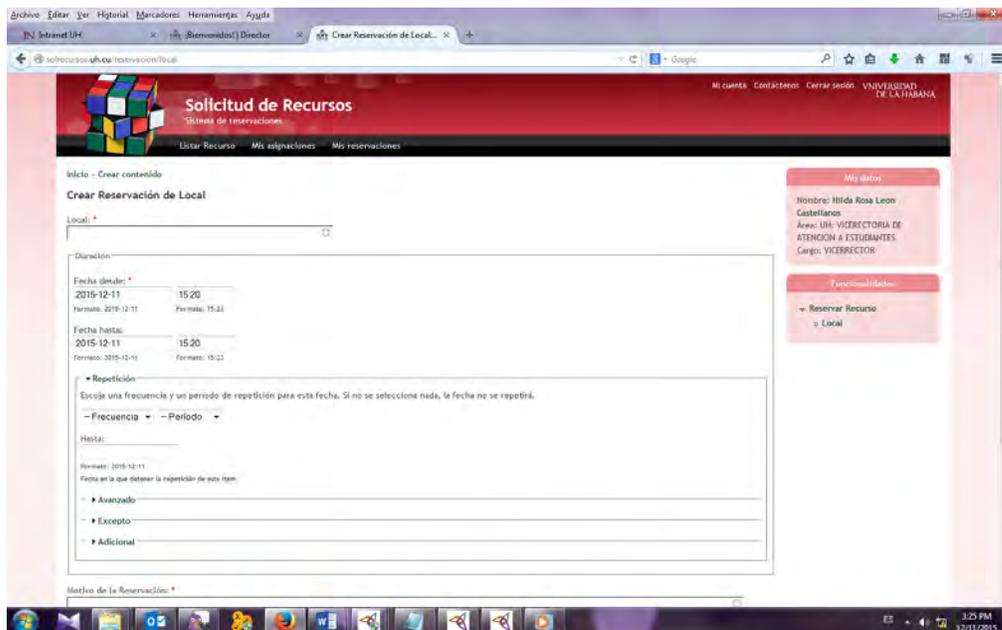


Fig. 5.61. Solicitud de un local: reservación en línea

Software: Drupal

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP. El sistema utiliza RSS

5.4.15. Sistema Análisis Periódico de Asistencia de la Universidad de La Habana APA-UH

Objetivo: El Sistema Análisis Periódico de Asistencia de la Universidad de La Habana (APA-UH) está confeccionado conjuntamente por la Dirección de Recursos Humanos y la Dirección de Informatización de la Universidad de la Habana. APA-UH es un sistema Web por ende puede ser visto desde cualquier computadora conectada a la red universitaria. Sólo pueden acceder a APA-UH las personas debidamente autorizadas por la dirección del área correspondiente. El sistema permite generar ausencias o llegadas tardes, ya sea diarias o por periodos de tiempo y emite los Análisis

Periódicos de Asistencia (APA) para los períodos de tiempo que se soliciten.
(Ver Fig 5.62)

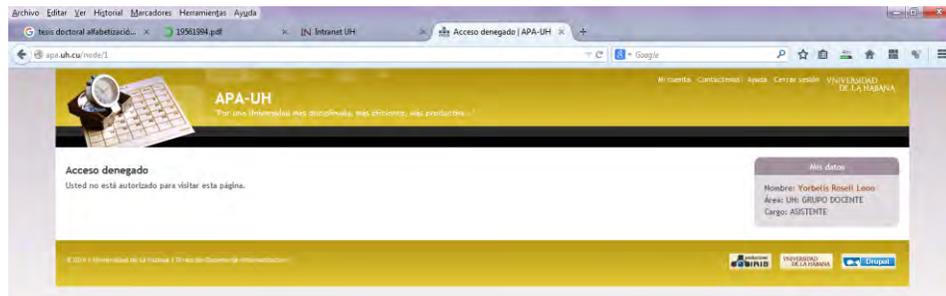


Fig. 5.62. Página de inicio del sitio APA-UH

Software: Drupal

Fuentes y datos: el sistema trabaja con datos primarios que se introducen en el mismo sistema los que son de tipo XML rpc. Además importa datos del Directorio Único en formato SOAP. El sistema utiliza RSS

5.4.16. Análisis del uso de CMS en la construcción de sistemas de gestión automatizados

En la UH hay un total de 14 sistemas centralizados para la gestión de información administrativa de la UH, de los cuales 12 están implementados en CMS eso significa un 85.7% del total. (Ver Fig. 5.63)

Drupal se impone como CMS predominante. La versión utilizada en todos los casos es la 6.0.

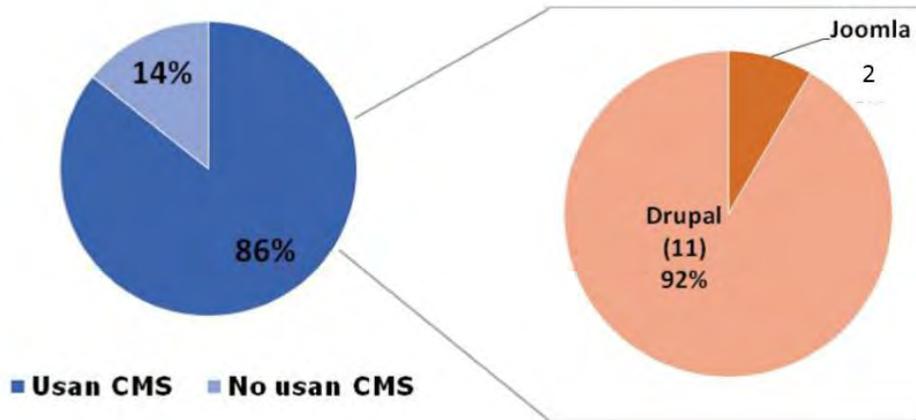


Fig. 5.63. Uso de CMS en la construcción de sistemas de gestión automatizados (elaboración propia)

Aunque con el diseño, implementación y uso de los diferentes sistemas la Universidad ha ganado mucho en la gestión de los procesos administrativos, la información aún se encuentra fragmentada y no hay manera de recuperarla o visualizarla de manera integrada y coherente.

Hay un extenso volumen de documentos en lenguaje natural expresado en lenguaje HTML y con escasa semántica, que provoca que las aplicaciones informáticas no siempre accedan de manera correcta a los contenidos, pues la mayor parte de estos documentos carecen de estructura, con excepción de elementos como título o etiquetas.

La autorganización de los documentos se da esencialmente a partir de enlaces y directorios y en menor grado por la asignación y uso de metadatos. Este último solo se da en algunos documentos vinculados a la publicación de información en la Intranet UH.

Puede valorarse que existe falta de normalización en las estructuras y los documentos, problemas de polisemia y sinonimia heredados del lenguaje natural.

Por otra parte, la autora reconoce que en la UH existen y desarrollan otros sistemas que no se limitan a la gestión de información administrativa, sino que van hacia la gestión del conocimiento gris, como tesis, investigaciones de todo tipo y a diversas escalas y alcances, etc. Para ellos, el esquema de

construcción de sitios web y los parámetros que hoy se siguen en dicha construcción, no funcionan. Para conseguir mayores funcionalidades se necesitaría crear un esquema que permita el procesamiento del lenguaje natural de manera que sea entendido por el sistema en un entorno tan heterogéneo como la UH con una adecuada y coherente estructuración de los documentos y desambiguación semántica, utilización de estadísticas y minería de datos, creación de sistemas de organización de conocimiento. Todo ello unido a la visión de crear y mejorar continuamente la interoperabilidad entre bases de datos o para los mecanismos de procesamiento por aplicaciones informáticas.

5.5. Análisis de la calidad de los datos gestionados por los sistemas de gestión automatizados en la UH

Con el objetivo de conocer el nivel de conocimiento y conformidad de los usuarios potenciales que utilizan los sistemas de gestión en la UH, se realizó un estudio basándose esencialmente en el criterio del Consejo Universitario reconocido en este contexto como el órgano de dirección máximo en la UH. Está compuesto por: 56 miembros, compuesto por Rector, Directores de Centros y Áreas de la UH, Decanos, Vicerrectores, Dirigentes de organizaciones políticas y de masa, asesores y otros miembros que son invitados permanentes.

5.5.1. Diseño de la encuesta

Para el estudio se establecieron un total de 31 indicadores distribuidos entre 3 de las variables (Ver Fig.5.64) definidas para el análisis de la calidad de los datos (fueron explicados en capítulo metodológico):

- Organización y Representación de la Información (ORI)
- Gestión de Información (GI)
- Búsqueda y recuperación de Información (BRI)

ORI	GI	BRI
<ul style="list-style-type: none"> •Exactitud •Compleitud •Consistencia •Objetividad •Interpretabilidad •Facilidad de comprensión •Consistencia representacional •Representación concisa 	<ul style="list-style-type: none"> •Tiempo de vida •Actualidad •Volatilidad •Veracidad •Credibilidad •Cantidad apropiada de datos •Accesibilidad •Unicidad •Seguridad •Utilidad •Preservación •Comodidad 	<ul style="list-style-type: none"> •Credibilidad •Precisión •Objetividad •Reputación •Valor agregado •Relevancia •Cantidad apropiada de datos •Interpretabilidad •Facilidad de comprensión •Accesibilidad •Comodidad

Fig. 5.64. Indicadores para la valoración de calidad de datos

Con ellos se preparó un ejercicio, desarrollados en torno a una encuesta (ver Anexo 2) diseñada sobre la base de preguntas cerradas para evitar ambigüedad en las respuestas, lo que facilitaría el procesamiento de las respuestas

5.5.2. Selección de la muestra

Para aplicación de la encuesta, se realizaron cálculos para conocer los posibles rangos a trabajar en las muestras (Cálculo del tamaño óptimo de la muestra):

- Margen de error máximo admitido: 10%
- Tamaño de la población: 68
- Tamaño para un nivel de confianza del 95%: 40
- Tamaño para un nivel de confianza del 97%: 43
- Tamaño para un nivel de confianza del 99%: 48

Para alcanzar el mayor nivel de fiabilidad posible y compromiso con la investigación por parte de los encuestados, se coordinaron 2 encuentros con el Consejo Universitario:

- En el primero de ellos (octubre/2015) se expuso el proyecto de investigación: objetivos, importancia, alcance, etc. Se repartió material con la exposición y los principales aspectos de interés que favorecerían la aplicación de la encuesta al utilizar los términos técnicos. Predominó un lenguaje simple y ejemplos prácticos.
- En el segundo se aplicó la encuesta (noviembre/2016). Con anterioridad se procedió a la lectura comentada de las preguntas y variables.

Se aplicó la encuesta a los 53 miembros y de ellos 1 de las encuestas tuvo que se invalidada por incoherencias en la respuestas. Desde el punto de vista estadístico se puede decir que:

- Tamaño de la población: 68
- Tamaño muestral realmente logrado: 52
- Error máximo para un nivel de confianza del 95%: 6.64%
- Error máximo para un nivel del confianza del 97%: 7.35%
- Error máximo para un nivel del confianza del 99%: 8.74%

De ello de deriva que, a pesar de no haberse obtenido un 100% de encuestas a procesar en función de la totalidad de la muestra, los valores son válidos para el análisis del pensamiento de la dirección de la comunidad de estudio.

5.5.3. Resultados del Análisis de la calidad de los datos

Dado el contexto de aplicación de la encuesta, todos los sujetos son graduados universitarios, en su mayoría con grado de Máster (16%) o Doctor (75%), con dominio sobre la política de impulso de la gestión de procesos administrativos en ambiente digital y la creación de intranets como parte de la cultura comunicacional de la institución.

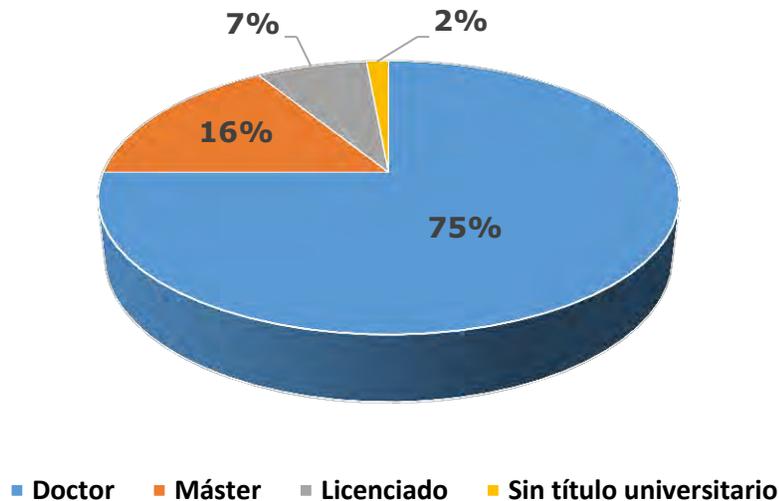


Fig. 5.65. Composición educacional del Consejo Universitario de la UH

Los resultados del estudio referente a lo que piensan los encuestados de las potencialidades de Drupal para el manejo de los datos se muestran a continuación (Ver tabla 5.1):

Tabla 5.1. Resultados de opinión sobre Drupal para el manejo de datos según los indicadores establecidos

Variables	Indicadores	Drupal				
		Baja ⁷²	Media ⁷³	Alta ⁷⁴	PROM ⁷⁵	Baja/Media/Alta
ORI	Exactitud	2	20	30	2.53846154	alta
	Complejidad	0	10	42	2.80769231	alta
	Consistencia	3	32	17	2.26923077	media
	Objetividad	9	8	35	2.5	alta
	Interpretabilidad	7	9	36	2.55769231	alta

⁷² Valor otorgado: 1 punto

⁷³ Valor otorgado: 2 puntos

⁷⁴ Valor otorgado: 3 puntos

⁷⁵ PROM: Es el promedio de las opiniones dado en cada indicador sobre el CMS en cuestión. Se calcula sumando la cantidad de personas que dieron determinada clasificación por el valor que corresponde a cada una y luego se divide por el total de personas encuestadas. Es decir: [(personas que consideran que es baja* 1)+(personas que consideran que es media*2)+(personas que consideran que es alta*3)]/total de personas

	Facilidad de comprensión	5	10	37	2.61538462	alta
	Consistencia representacional	2	11	39	2.71153846	alta
	Representación concisa	10	15	27	2.32692308	media
GI	Tiempo de vida	0	2	50	2.96153846	alta
	Actualidad	3	10	39	2.69230769	alta
	Volatilidad	7	9	36	2.55769231	alta
	Veracidad	1	3	48	2.90384615	alta
	Credibilidad	0	0	52	3	alta
	Cantidad apropiada de datos	2	4	46	2.84615385	alta
	Accesibilidad	0	0	52	3	alta
	Unicidad	0	5	47	2.90384615	alta
	Seguridad	12	6	34	2.42307692	media
	Utilidad	2	1	49	2.90384615	alta
	Preservación	7	0	45	2.73076923	alta
	Comodidad	1	1	50	2.94230769	alta
BRI	Credibilidad	1	1	50	2.94230769	alta
	Precisión	2	1	49	2.90384615	alta
	Objetividad	1	1	50	2.94230769	alta
	Reputación	1	6	45	2.84615385	alta
	Valor agregado	1	8	43	2.80769231	alta
	Relevancia	6	1	44	2.69230769	alta
	Cantidad apropiada de datos	3	6	43	2.76923077	alta
	Interpretabilidad	0	2	50	2.96153846	alta
	Facilidad de comprensión	5	7	40	2.67307692	alta
	Accesibilidad	2	2	48	2.88461538	alta
	Comodidad	2	6	44	2.80769231	alta

En cuanto a lo que piensan los encuestados de las potencialidades de Joomla para el manejo de los datos se tiene: (Ver además, Anexo 3)

Tabla 5.2. Resultados de opinión sobre Joomla para el manejo de datos según los indicadores establecidos

Variables	Indicadores	Joomla				
		Baja	Media	Alta	PROM	Baja/ Media/ Alta
ORI	Exactitud	30	20	2	1.46153846	baja
	Compleitud	34	10	18	2.07692308	media
	Consistencia	40	10	2	1.26923077	baja
	Objetividad	45	5	2	1.17307692	baja
	Interpretabilidad	23	13	16	1.86538462	media
	Facilidad de comprensión	18	2	32	2.26923077	media
	Consistencia representacional	6	10	36	2.57692308	alta
	Representación concisa	4	6	42	2.73076923	alta
GI	Tiempo de vida	1	2	49	2.92307692	alta
	Actualidad	12	11	29	2.32692308	media
	Volatilidad	10	12	32	2.5	alta
	Veracidad	18	3	31	2.25	media
	Credibilidad	0	0	52	3	alta
	Cantidad apropiada de datos	0	25	27	2.51923077	alta
	Accesibilidad	0	0	52	3	alta
	Unicidad	2	23	25	2.36538462	media
	Seguridad	2	3	47	2.86538462	alta
	Utilidad	3	3	46	2.82692308	alta
	Preservación	10	8	34	2.46153846	media
	Comodidad	2	5	46	2.88461538	alta
BRI	Credibilidad	2	1	49	2.90384615	alta
	Precisión	3	1	48	2.86538462	alta
	Objetividad	8	9	35	2.51923077	alta
	Reputación	2	6	44	2.80769231	alta
	Valor agregado	6	9	37	2.59615385	alta
	Relevancia	4	5	43	2.75	alta
	Cantidad apropiada de	6	1	44	2.69230769	alta

	datos					
	Interpretabilidad	3	1	48	2.86538462	alta
	Facilidad de comprensión	10	2	40	2.57692308	alta
	Accesibilidad	2	3	47	2.86538462	alta
	Comodidad	5	6	41	2.69230769	alta

Al comparar los datos obtenidos se puede apreciar que se tiene en mejor concepción el uso de Drupal para el manejo de datos en casi todos los aspectos (Ver tabla 5.3 y Fig. 5.66. Ver además, Anexo 3)

Tabla 5.3. Tabla comparativa entre Drupal y Joomla para el manejo de datos según los indicadores establecidos

Variables	Indicadores	Drupal Prom	Joomla Prom	Mejor Drupal	Mejor Joomla
ORI	Exactitud	2.53846154	1.46153846	X	
	Compleitud	2.80769231	2.07692308	X	
	Consistencia	2.26923077	1.26923077	X	
	Objetividad	2.5	1.17307692	X	
	Interpretabilidad	2.55769231	1.86538462	X	
	Facilidad de comprensión	2.61538462	2.26923077	X	
	Consistencia representacional	2.71153846	2.57692308	X	
	Representación concisa	2.32692308	2.73076923		X
GI	Tiempo de vida	2.96153846	2.92307692	Similares	
	Actualidad	2.69230769	2.32692308	X	
	Volatilidad	2.55769231	2.5	Similares	
	Veracidad	2.90384615	2.25	X	
	Credibilidad	3	3	Similares	
	Cantidad apropiada de datos	2.84615385	2.51923077	X	
	Accesibilidad	3	3	Similares	
	Unicidad	2.90384615	2.36538462	X	

	Seguridad	2.42307692	2.86538462		X
	Utilidad	2.90384615	2.82692308	X	
	Preservación	2.73076923	2.46153846	X	
	Comodidad	2.94230769	2.88461538	X	
BRI	Credibilidad	2.94230769	2.90384615	Similares	
	Precisión	2.90384615	2.86538462	Similares	
	Objetividad	2.94230769	2.51923077	X	
	Reputación	2.84615385	2.80769231	Similares	
	Valor agregado	2.80769231	2.59615385	X	
	Relevancia	2.69230769	2.75	Similares	
	Cantidad apropiada de datos	2.76923077	2.69230769	X	
	Interpretabilidad	2.96153846	2.86538462	X	
	Facilidad de comprensión	2.67307692	2.57692308	X	
	Accesibilidad	2.88461538	2.86538462	Similares	
	Comodidad	2.80769231	2.69230769	X	

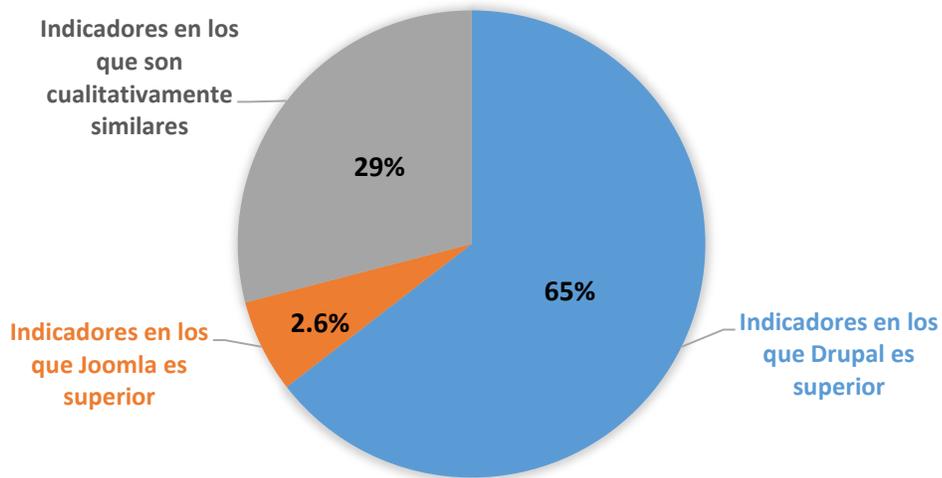


Fig. 5.66. Comportamiento de los indicadores recogidos

Los resultados obtenidos, no son más que el reflejo de la tendencia que se ha dado en el contexto de la Universidad de La Habana en los últimos cinco años. Las apreciaciones de los dirigentes en cuanto al manejo de datos, responden al interés por mejorar la gestión de la gestión universitaria,

hacerla cada vez más transparente, ágil y confiable. Es por ello que desde la dirección universitaria, especialmente del equipo asesor que lleva los temas desde el Rectorado, la Dirección de información y la Dirección de Informatización , se ha trabajado en la alfabetización informacional sobre los temas que atañen el desarrollo del proceso e implementación de los sistemas.

5.6. CMS más utilizados en la Universidad de La Habana

El objetivo del presente epígrafe es estudiar las características genéricas fundamentales y características semánticas de Drupal y Joomla, al ser identificados como los CMS más utilizados para la construcción de sitios en la UH.

5.6.1. Drupal

Drupal es un CMS gratuito y de código abierto distribuido bajo la licencia GPL⁷⁶. Surgió en 2001 y es mantenido y desarrollado por una comunidad de miles de usuarios y desarrolladores. Permite publicar, gestionar y organizar fácilmente el contenido en sitios web, incluyendo portales comunitarios web, sitios de discusión, sitios web, aplicaciones de Intranet Corporativa, sitios web personales o blogs, aplicaciones de comercio electrónico, directorios de recursos, sitios de redes sociales (CMS-Matrix, 2015).

5.6.1.1. Características

Drupales una aplicación de código abierto. Diseñado para entornos PHP. Dispone de una capa de abstracción de base de datos para trabajar con diversas bases de datos: MySQL, PostgreSQL, SQLite, MS-SQL-Server, etc. Pensado para funcionar sobre la base de Apache como servidor web y cualquier sistema operativo.

Orientado a crear proyectos colaborativos y comunidades de usuarios. Posee gestión de usuarios y potente sistema de permisos basado en roles (grupos de usuarios). Gestión de contenidos avanzada: control de versiones,

⁷⁶ GPL: General Public License

tipos de contenido, uso avanzado de taxonomías. Arquitectura modular que permite ampliar y personalizar funcionalidades

Su curva de aprendizaje más lenta que otros CMS como Wordpress o Joomla

Como CMS Drupal posee:

- **Elementos de seguridad:** Proceso de aprobación de contenidos antes de ser publicados, verificación por e-mail para el registro del usuario, registro por parte de la red a cada una de las consultas de los recursos, autorizados o no (LoginHistory), mecanismos para la identificación de usuarios común y centralizado (PluggableAuthentication), gestión de sesiones para mantener al tanto de las actividades a lo largo de una sesión (Session Management), protocolo de seguridad (SSL Compatible), creación del registro que se genera con el rastro de las actividades que se realizan en el sitio de manera cronológica (AuditTrail), método para controlar y limitar al acceso de los usuarios dentro del sitio (Granular Privileges).

- **Materiales de apoyo:** Manuales, demos, comunidad de desarrollo, ayuda on-line, interfaz de programación de aplicación de uso general para evitar a los programadores trabajar todo desde el principio (Pluggable API), servicio de alojamiento ofrecido por empresas (Professional Hosting), posibilidad de contratar servicios (Professional Services), los usuarios pueden compartir información a través del foro público (PublicForum), componentes de terceros o aplicaciones para la gestión de contenidos que pueden ser añadidos según las necesidades (Third-PartyDevelopers), eventos que se celebran de manera anual para compartir información con carácter técnico entre desarrolladores y usuarios (UsersConference), lista de direcciones de correo electrónico utilizadas para enviar mensaje a un grupo de personas (PublicMailingList).

- **Facilidades de uso:** URL amigables que resultan mejor indexadas en los motores de búsqueda de internet (FriendlyURLs), para generar páginas web de forma dinámica en el servidor (Server Page Language), contiene plantillas prediseñadas (TemplateLanguage). Permite adicionar módulos y aplicaciones para modificar las dimensiones de las imágenes de manera

individual o en conjunto (ImageResizing), lenguaje de programación en la que la mayoría de los códigos está compuesto por macros (Macro Language), transferencia de información en grandes cantidades para compartir con los demás usuarios (MassUpload), permite crear prototipos para visualizar los sitios antes de crearlos (Prototyping), y WYSIWYG Editor como herramienta que se aplica a los procesadores de texto.

- **Interoperabilidad:** Sindicación de contenidos (RSS) como sistema de alertas para conocer cuándo se ha hecho actualización de contenidos sin tener que acceder a ellas; UTF-8 Support para la codificación de caracteres.

- **Flexibilidad:** Interfaz para el intercambio de datos estándar (CGI-modeSupport), perfiles de usuarios extensibles que permite la personalización de determinados espacios sin afectar la configuración general (Extensible UserProfiles), Interface Localization, Metadata como herramienta para la recuperación de información, URL Rewriting

- **Gestión:** administración de fondos de inversión como ayuda a los clientes (Asset Management), herramienta para el diseño, herramientas para administración del sitio (InlineAdministration, Online Administration), sección de un sitio enfocada a un tema particular (Sub-sites / Roots), conjunto de íconos gráficos para el uso de los usuarios (Themes / Skins), programa que permite ordenar y gestionar estadísticas de la página web (Web Statistics), diseño o estilos de la página web con la utilización de plantillas (Web-based Style/Template Management).

- **Aplicaciones:** Blog, preguntas frecuentes (FAQ Management), Link Management, correo, sondeos y encuestas, Product Management, Discussion/ Forum, pero puede incorporar módulos y aplicaciones para el desarrollo de Chat, clasificados, datos de entrada, calendario de eventos, File Distribution, Groupware, Guest Book, HelpDesk / Bug Reporting, boletín de noticias, Galería de fotos, seguimiento de proyectos, mapa del sitio, contribuciones de usuario, Wiki, gestión de contactos.

5.6.1.2. Evaluación de características semánticas

Según avanzan las aplicaciones de Drupal, han ido buscando mayores potencialidades, como en cualquier proceso de desarrollo sobre la base de la construcción colectiva.

En la versión Drupal 7 se aprecian cambios en la interfaz de administración que han mejorado su usabilidad, mejoras en el soporte de base datos: posibilidad usar SQLite, mayor escalabilidad y consistencia en API de base de datos, así como la incorporación de CCK (tipos de contenido) en el núcleo del sistema.

En las últimas versiones de Drupal, es posible utilizar módulos como RDF extensions, Microdata y Schema.org (ver fig. 5. 67a la fig. 5. 73) relacionados con la publicación semántica. Estos módulos permiten anotar contenidos de Drupal, de manera que esa información se publique de forma estructurada tanto en el marcado (utilizando RDFa o Microdata) o en crudo (RDF). En general, estas extensiones permiten anotar el contenido de forma manual haciendo uso de herramientas libres como Apache Stanbol, Apache Marmotta y Apache Solr junto con los módulos anteriores, se puede potenciar la publicación semántica en Drupal (Pastor Sánchez, 2011a).

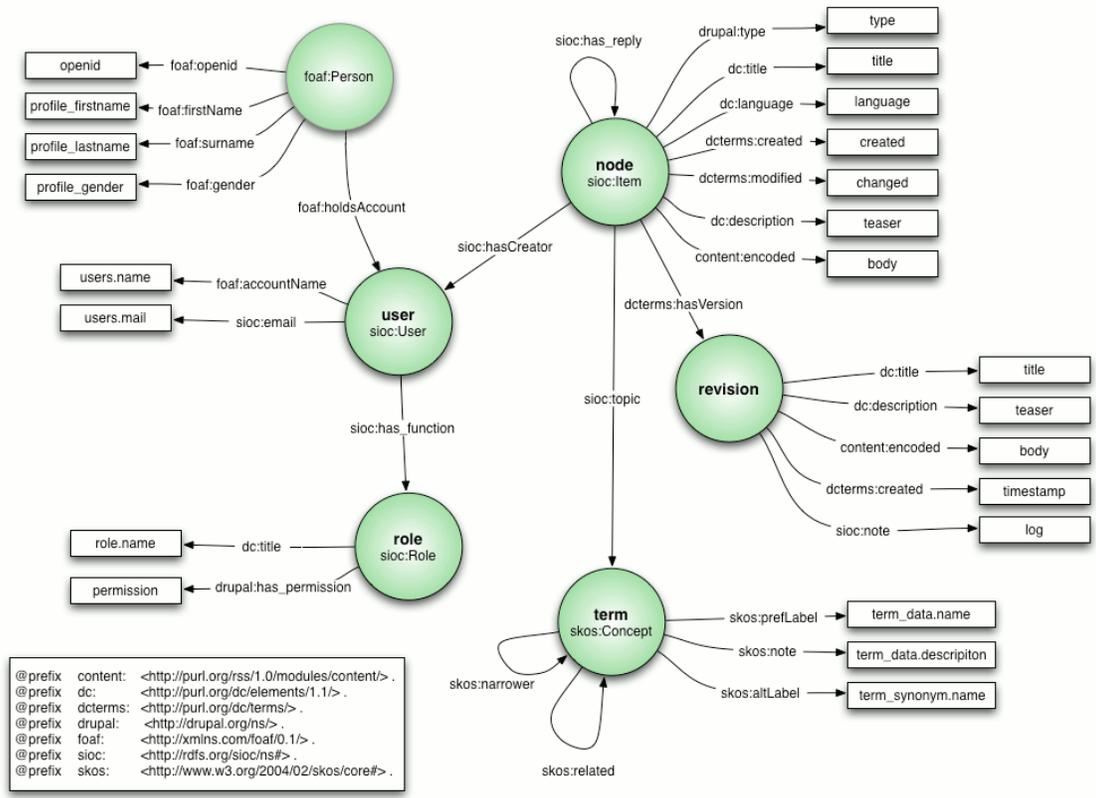


Fig. 5.67. Modelo RDF en Drupal 7: Representación de nodos, usuarios, roles, temas y versiones con RDF (Koller, 2008)



Fig. 5.68. Módulo básico del soporte RDF en Drupal 7 (Pastor Sánchez, 2011a)



Fig. 5.69. Módulos que amplían el soporte nativo RDF en Drupal 7 (Pastor Sánchez, 2011a)

RDF publishing settings

Manage the namespaces and associated prefixes used by the site. Prefixes allow URIs to be shortened in the form of CURIEs (Compact URIs). For example, the CURIE `dc:title` represents the URI `http://purl.org/dc/terms/title`.

PREFIX	NAMESPACE
content	http://purl.org/rss/1.0/modules/content/
dbpedia	http://dbpedia.org/resource/
dc	http://purl.org/dc/terms/
foaf	http://xmlns.com/foaf/0.1/
frbr	http://purl.org/vocab/frbr/core#
og	http://ogp.me/#
rdfs	http://www.w3.org/2000/01/rdf-schema#
sioc	http://rdfs.org/sioc/ns#
sioc:ct	http://rdfs.org/sioc/types#
skos	http://www.w3.org/2004/02/skos/core#
xsd	http://www.w3.org/2001/XMLSchema#
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rss	http://purl.org/rss/1.0/

Prefijo *

Choose a prefix for this namespace, e.g. dc, foaf, sioc. This prefix will be used as an abbreviation for the namespace URI.

Namespace URI *

Enter the URI of the namespace. Make sure it ends with either / or #

Fig. 5.70. Vocabularios RDF externos en Drupal 7: Incorporación de vocabularios mediante External RDF Vocabulary Importer (Pastor Sánchez, 2011a)

Referencias Bibliográficas

Nombre de sistema: `referencias_bibliograficas` [Editar]

El nombre legible para humanos de este tipo de contenido. Este texto será mostrado como parte del listado en la página *Agregar nuevo contenido*. Es recomendable que este nombre comience con una letra mayúscula y contenga solamente letras, números y espacios. Este nombre debe ser único.

Descripción
 Tipo de contenido para la gestión de Referencias Bibliográficas

Describir este tipo de contenido. Este texto es lo que se muestra en la página *Añadir nuevo contenido*.

Opciones del formulario de envío
 Título

Opciones de publicación
 Publicado - Colocado en la página principal

Opciones de presentación
 Mostrar información del autor y fecha.

Opciones de comentarios
 Abrir - Hilos de comentarios - 50 comentarios por página

Opciones del menú

RDF Settings

RDF Type
 dc:BibliographicResource

Enter a comma-separated list of classes for this bundle using CURIE syntax. For example: `sioc:item, foaf:Document`

RDF Predicates

Enter a comma-separated list of predicates for title using CURIE syntax. For example: `foaf:familyName, foaf:lastName`

Attribute Type
 property

For fields containing literals—things such as plain text, html, numbers—use the `property` attribute. For fields containing references to other things—URIs and node references, for example—use the `rel` or `rev` attribute.

Datatype

Enter the datatype for title using CURIE syntax. For a list of common datatypes, see: [XML Schema datatypes](#).

Fig. 5.71. Tipos de Contenidos y RDF en Drupal 7: Definición de objetos RDF a partir de tipos de contenidos en Drupal (Pastor Sánchez, 2011a)

Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives

Ver Editar RDF

Autor / Autores:
 Juan Antonio Pastor Sánchez
 Francisco Javier Martínez Méndez
 José Vicente Rodríguez Muñoz

```

<div class="field-item" property="dc:creator">Juan Antonio Pastor Sánchez</div>
<div class="field-item" property="dc:creator">Francisco Javier Martínez Méndez</div>
<div class="field-item" property="dc:creator">José Vicente Rodríguez Muñoz</div>
    
```

Drupal renderiza el código XHTML del documento insertando atributos RDFa a partir del mapeado definido

Mediante técnicas de parseado o servicios de extracción RDFa como <http://www.w3.org/2007/09/pyRDFa> es posible extraer sentencias RDF del código XHTML del documento

```

<dc:creator xml:lang="es">Juan Antonio Pastor Sánchez</dc:creator>
<dc:creator xml:lang="es">José Vicente Rodríguez Muñoz</dc:creator>
<dc:creator xml:lang="es">Francisco Javier Martínez Méndez</dc:creator>
    
```

This paper presents an analysis of the Simple Knowledge Organization System (SKOS) compared with other alternatives for thesaurus representation in the Semantic Web. Based on functional and structural changes of thesauri, provides an overview of the current context in which lexical paradigm is abandoned in favour of the conceptual paradigm. Likewise are described briefly various initiatives for the representation of thesauri using RDF vocabularies and hence for application to the Semantic Web, with particular attention to SKOS. This brief comparison will allow us to know some highlights of these proposals to raise those transcendental aspects in a model representing thesauri. SKOS includes the main proposal of this model, which organizes the concepts into diagrams and libraries. The notions of descriptors and non-descriptors terms are replaced by the association to the concepts of preferred or alternative labels and may be defined hierarchical or associative relationships between concepts. SKOS also provides for the establishment of correspondence relations between concepts belonging to different conceptual schemes. In accordance with the extent to which this series of requirements is fulfilled, we propose and defend the SKOS option as the best alternative for the development of multiple applications focusing on the use of thesauri in Web-based information services and systems, highlighting the advantages of this model over others proposed and its consideration of the user perspective in management, search and information query operations.

Año de publicación:
2009

Referencia Bibliográfica (estilo APA):
Pastor, J.A., Martínez, F.J. & Rodríguez J.V. (2009). "Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives". Information Research, 14(4) paper 422. [Available at <http://InformationR.net/ir/14-4/>]

Datos de edición:
Information Research, 14(4) paper 422

Ubicación:
<http://informatics.net/ir/14-4/paper422.html>

Fig. 5.72. Soporte RDFa en Drupal 7: Inserción de RDFa en el código XHTML (Pastor Sánchez, 2011a)

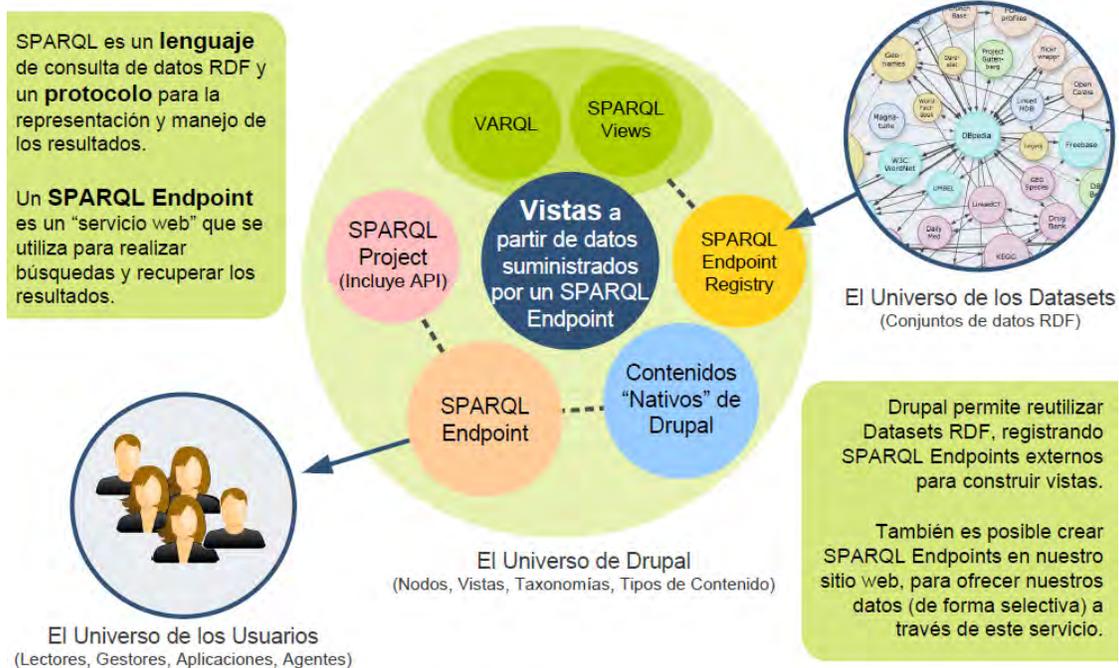


Fig. 5.73. Soporte SPARQL en Drupal 7: El ecosistema SPARQL en Drupal (Pastor Sánchez, 2011a)

A modo de resumen, después de consultar algunos especialistas (ver anexo 1) se presenta un análisis sobre las características semánticas de Drupal

Tabla 5.4. Resultado de la encuesta realizada a especialistas sobre Drupal

CMS	Indicadores	Evaluación
Drupal	Soporte nativo para la gestión de vocabularios y ontologías	Drupal solo es capaz de manejar datos de escasos vocabularios ontológicos, entre ellos: DublinCore, Foaf, Sioc y SKOS. Es inoperante cuando se intenta describir ontologías específicas que combinen muchas propiedades en los datos, por razón muchas necesidades de usuarios con altos niveles de mapificación de datos se ven trucadas al no poderse manejar ontologías específicas como: VIVO, bibo, LOM, geo-ontologías, etc. Se conoce que dentro del CORE de Drupal existen módulos que interpretan ontologías, tal es el caso de OSF desarrollado en el 2013 ha mejorado la interacción con otros formatos ontológicos, pero deja detrás las capacidades de inferencia semánticas compleja y los casos de manejo de lógicas descriptivas complejas dentro de las ontologías que son manejadas con procesos de implicación simple y taxonomías muy desarrolladas para suplir las deficiencias de este módulo.
	Asociación entre los contenidos estructurados del portal y los vocabularios RDF gestionados	El Sistema es capaz de asociar datos y contenidos en formatos RDF y RDFa capaces de mostrar datos coherentes con un buen nivel semántico, se logran realizar mapeos, pero solo de forma sintáctica, es decir los datos se mueven de forma estructural y no semántica, pues el proceso de asociación requiere técnicas de alineación de ontologías que buscan la similitud entre los vocabularios.

		Posibilidad de asociar cada tipo estructurado a una o varias clases RDF y mapear parte de los campos del tipo estructurado a propiedades RDF y RDFa.
	Adquisición de módulos	Tiene todas las facilidades de modularidad, integra módulos para lograr sus funciones básicas en el manejo de datos RDF, al presentar los módulos siguientes: RDF Indexer, RDF, Entity RDF, RDF UI, RDF extensions, Simple RDF, RDF CCK, RDF Mapin, RDF Hub, RDF SparqlEndpoint, RDF SPARQL Proxi, JSO LDF, entre otros.
	Asociaciones RDF	Posibilita de forma eficiente las asociaciones entre datos RDF, pero si los datos no están en este formato no posee capacidad de migración. Por otra parte solo consigue alinear datos sintácticamente , de forma completamente semántica
	SparQL Endpoints y Linked Data	Con el módulo OSF se da la posibilidad de manejar datos abiertos o datos enlazados y de realizar consultas a través de SPARQL Endpoint, pero solo sobre limitados recursos de información que exponen sus datos, en ocasiones las apis para manejar estos datos están poco desarrolladas y son proyectos en desarrollo.

5.6.2. Joomla

Joomla es una aplicación de código abierto, construido en PHP y MySQL como base de datos y está disponible bajo licencia GPL. Según Mooney y Baenziger (2007), Joomla es descendiente de otro CMS: Mambo. La primera versión fue publicada en el 2005.

Es una versión estable, con un código completamente reescrito que incorpora notables mejoras en el área de seguridad, administración y cumplimiento con estándares W3C y construido bajo PHP5. Por lo general, es la que se está implementando en las universidades estudiadas.

Ofrece una arquitectura de aplicación que permite a los desarrolladores crear extensiones que amplían la capacidad de Joomla en múltiples direcciones, identificados como los llamados "addon".

5.6.2.1. Características técnicas

Entre las características más valoradas por los especialistas que lo seleccionan se refieren a que Joomla es utilizado en todo el mundo para poder crear sitios Web de todas las formas y tamaños, manifestando la capacidad de adaptar el sistema a las necesidades propias (flexibilidad) para la implantación de:

- Sitios Web corporativos o portales
- Intranets y extranets corporativas
- Revistas en línea, periódicos y publicaciones
- Sitios Web de pequeñas empresas
- Sitios Web sin fines de lucro
- Portales basados en la comunidad
- Páginas web personales o familiares

Como CMS, Joomla posee:

Elementos de seguridad: Proceso de aprobación de contenidos antes de ser publicados, verificación por e-mail para el registro del usuario, servicio de directorio para buscar información en un entorno de red a partir del protocolo de acceso unificado (LDAP Authentication), registro por parte de la red a cada una de las consultas de los recursos, autorizados o no (LoginHistory), mecanismos para la identificación de usuarios común y centralizado (PluggableAuthentication), gestión de sesiones para mantenerse al tanto de las actividades a lo largo de una sesión (Session Management), protocolo de seguridad (SSL Compatible, SSL Logins, SSL Pages).

Materiales de apoyo: Manuales, demos, comunidad de desarrollo, ayuda on-line, interfaz de programación de aplicación de uso general para evitar a los

programadores trabajar todo desde el principio (Pluggable API), servicio de alojamiento ofrecido por empresas (Professional Hosting), posibilidad de contratar servicios (Professional Services), los usuarios pueden compartir información a través del foro público (PublicForum), componentes de terceros o aplicaciones para la gestión de contenidos que pueden ser añadidos según las necesidades (Third-PartyDevelopers), eventos que se celebran de manera anual para compartir información con carácter técnico entre desarrolladores y usuarios (UsersConference).

Facilidades de uso: URL amigables que resultan mejor indexadas en los motores de búsqueda de internet (FriendlyURLs), permite modificar las dimensiones de las imágenes de manera individual o en conjunto (ImageResizing), lenguaje de programación en la que la mayoría de los códigos está compuesto por macros (Macro Language), transferencia de información en grandes cantidades para compartir con los demás usuarios (MassUpload), permite crear prototipos para visualizar los sitios antes de crearlos (Prototyping), permite generar páginas Web de forma dinámica en el servidor (Server Page Language), contiene plantillas prediseñadas (TemplateLanguage), existencia de diferentes niveles de interacción del usuario con el sistema (UI Levels), cuenta con la herramienta WYSIWYG Editor que se aplica a los procesadores de texto.

Interoperabilidad: Sindicación de contenidos(RSS) como sistema de alertas para conocer cuándo se ha hecho actualización de contenidos sin tener que acceder a ellas; FTP Support; UTF-8 Support para la codificación de caracteres.

Flexibilidad: Interfaz para el intercambio de datos estándar (CGI-modeSupport), reutilización del contenido (Content Reuse), perfiles de usuarios extensibles que permite la personalización de determinados espacios sin afectar la configuración general (Extensible UserProfiles), Interface Localization, Metadata como herramienta para la recuperación de información, URL Rewriting

Gestión: Gestión de publicidad a través de los banners (Advertising Management), administración de fondos de inversión como ayuda a los clientes (Asset Management), herramienta para el diseño, ejecución y evaluación del problema de planificación de los contenidos del sitio Web (Content Scheduling), herramientas para administración del sitio (InlineAdministration, Online Administration), sección de un sitio enfocada a un tema particular (Sub-sites / Roots), conjunto de íconos gráficos para el uso de los usuarios (Themes / Skins), papelera (Trash), programa que permite ordenar y gestionar estadísticas de la página web (Web Statistics), diseño o estilos de la página web con la utilización de plantillas (Web-based Style/Template Management)

Aplicaciones: Blog, gestión de contactos, preguntas frecuentes (FAQ Management), Link Management, correo, sondeos y encuestas, Product Management, pero puede incorporar módulos y aplicaciones para el desarrollo de Chat, clasificados, datos de entrada, DatabaseReports, Discussion / Forum, Document Management, calendario de eventos, Expense Reports, File Distribution, herramientas gráficas para la representación de datos, Groupware, Guest Book, HelpDesk / Bug Reporting, boletín de noticias, Galería de fotos, seguimiento de proyectos, mapa del sitio, contribuciones de usuario, Wiki.

5.6.2.2. Evaluación características semánticas

A modo de resumen, en la tabla 5.5, después de consultar algunos especialistas (ver anexo 1) se presenta un análisis sobre las características semánticas de Joomla.

Tabla 5.5. Resultado de la encuesta realizada a especialistas sobre Joomla

Variable	Indicadores	Definición
Joomla	Soporte nativo para la gestión de vocabularios y ontologías	Este CMS es incapaz de gestionar datos en RDF o RDFa, su estructura no están descrita para usar datos semánticos
	Asociación entre	No ofrece la posibilidad de asociar cada tipo

	los contenidos estructurados del portal y los vocabularios RDF gestionados	estructurado a una o varias clases RDF y mapear parte de los campos del tipo estructurado a propiedades RDF y RDFa.
	Adquisición de módulos	No ofrece posibilidades de extensión e integración a través de la vinculación a módulos extras para extender la funcionalidad básica del CMS con posibilidades de acciones semánticas
	Asociaciones RDF	No posibilita la publicación de las asociaciones RDF creadas en formatos semánticos
	SparQL Endpoints y Linked Data	No manejan datos abiertos o datos enlazados y de realizar consultas a través de SPARQL Endpoint

5.7. Referencias bibliográficas

- AGUILERA, F. 2016. Entrevista a especialista de la Dirección de Formateización [27 de enero de 2016]. Universidad de La Habana. La Habana, Cuba
- CENTRO DE ESTUDIOS DE ADMINISTRACIÓN PÚBLICA. 2015. Quiénes somos [Online. Available: <http://www.ceap.uh.cu/> [Accessed 13 de enero de 2016].
- CENTRO DE ESTUDIOS DE LA ECONOMÍA CUBANA DE LA UNIVERSIDAD DE LA HABANA. 2016. ¿Quiénes somos? [Online]. Available: http://www.ceec.uh.cu/que_es_el_ceec [Accessed 13 de enero de 2016].
- CENTRO DE ESTUDIOS DEMOGRÁFICOS DE LA UNIVERSIDAD DE LA HABANA. 2012. Página de inicio de la intranet [Online]. Available: <http://intranet.cedem.uh.cu/> [Accessed 15 de enero de 2016].
- CENTRO DE ESTUDIOS DEMOGRÁFICOS DE LA UNIVERSIDAD DE LA HABANA. 2012. *Página de inicio de la intranet* [Online]. Available: <http://intranet.cedem.uh.cu/> [Accessed 15 enero 2016].
- CENTRO DE ESTUDIOS PARA EL PERFECCIONAMIENTO DE LA EDUCACIÓN SUPERIOR (CEPES) DE LA UNIVERSIDAD DE LA HABANA. 2012. ¿Quiénes somos? [Online]. La Habana. Available: <http://cepes.uh.cu> [Accessed 13 de enero de 2016].
- CENTRO DE INVESTIGACIONES MARINAS. 2015. Informe de Autoevaluación Institucional (2010 - Junio 2015).
- FACULTAD LATINOAMERICANA DE CIENCIAS SOCIALES (FLACSO). 2014. *Quiénes somos* [Online]. Available: <http://www.flacso.uh.cu/index.php/quienes-somos> [Accessed 13 de enero de 2016].
- INSTITUTO DE FARMACIA Y ALIMENTOS. 2012. *Intranet* [Online]. La Habana. Available: <http://intranet.ifal.uh.cu/> [Accessed 15 enero 2016].
- IRIS. 2011. *Intranet UH: enlace a Facultades* [Online]. La Habana: Grupo Producciones IRIS. Available: <http://intranet.uh.cu/enlaces/facultades> [Accessed enero 12 2016].
- KOLLER, A. 2008. Semantic Web and Drupal. *The Semantic Puzzle*. 13 marzo 2008 ed.
- PASTOR SÁNCHEZ, J. A. 2011. Drupal 7: Web Semántica al alcance de todos. *Fesabid 2011*.
- ROSELL LEÓN, Y. 2009. *Impacto de los Sistemas Gestores de Contenido (CMS) en Centros de Educación Superior de Ciudad de La Habana*. Diploma de Estudios Avanzados del Doctorado en Documentación e Información Científica, Universidad de La Habana (Cuba) – Universidad de Granada (España).

RUIZ JHONES, A. 2016. Entrevista a Directora de Informatización de la Universidad de La Habana[27 de enero de 2016]. La Habana, Cuba.

UNIVERSIDAD DE LAHABANA. 2010. *Intranet de la Universidad de La Habana* [Online]. La Habana: Producciones IRIS. Available: <http://intranet.uh.cu/universidad-de-la-habana/mision-vision-y-objetivos> [Accessed 20 de septiembre de 2015].

CAPÍTULO VI

PROPUESTA MODÉLICA DEL CMS SEMÁNTICO UH-WEB PARA LA CREACIÓN DE SITIOS WEB EN LA UNIVERSIDAD DE LA HABANA

CAPÍTULO VI: Propuesta modélica del CMS UH-WEB para la creación de sitios web en la Universidad de La Habana.

La Universidad de La Habana (UH) en los últimos cinco años ha realizado varias transformaciones en su sistema de comunicación de manera general y de gestión de manera particular, dándole mayor protagonismo al uso de sistemas de gestión por vía digital desde las diferentes áreas de acción universitaria: docencia, investigación, actividades extensionistas a la comunidad y los procesos administrativos internos que todo ello genera. Dicha situación se ha visto reflejada en las arquitecturas de sus intranets que, en su mayoría, han evolucionado hacia un modelo centrado en el usuario y la gestión por procesos de trabajo.

Por todo ello, se impone la necesidad de ofrecer una herramienta que permita a estudiantes, profesores, investigadores y trabajadores una manera de encontrar respuestas a sus preguntas de forma más rápida y sencilla, con información precisa.

Al insistir en el trabajo de un sistema de recursos digitales accesibles desde una web para la UH de más semántica, se pueden obtener soluciones más rápidas y coherentes a problemas habituales en la búsqueda y recuperación de información por medio de la utilización de una infraestructura común, mediante la cual, sea posible compartir, procesar y transferir información de forma sencilla.

La propuesta de diseño de un modelo de CMS semántico para la construcción de los sitios web en la UH, tributa al desarrollo y uso de la Web Semántica en este contexto, ayudando a resolver problemas como la sobrecarga de información y heterogeneidad de fuentes de información con el derivado problema de interoperabilidad que, en este contexto, resultan cada vez más agobiantes dada la complejidad de su estructura y riqueza temática en la que se desenvuelve. El desarrollo y futura implementación de la nueva propuesta permitirá a los

usuarios delegar tareas al software, pues gracias a la semántica en la Web, será capaz de procesar su contenido, razonar con este, combinarlo y realizar deducciones lógicas para resolver problemas cotidianos automáticamente.

6.1. Diseño de la propuesta de CMS Semántico para la UH: UH-WEB

En el epígrafe se exhibe un modelo personalizado de CMS semántico para la UH. La propuesta se construye desde la óptica de los procesos de producción de información en el área de la Ciencias de la Información.

Del modelo se presentan:

- Principios y requerimientos
- Objetivos
- Requisitos de diseño
- Arquitectura
- Ontología y lenguajes ontológicos
- Esquema de anotación semántica
- Recuperación de información

En conformidad con lo expuesto durante la investigación, la autora propone el diseño de un CMS que responde al problema científico de la investigación y está dirigido a ayudar y orientar a desarrolladores, expertos en información y documentación, a las autoridades y factores que dirigen el proceso de informatización en la UH, así como a los investigadores e interesados en el tema. Como es característico a todo diseño de CMS se le definen objetivos, principios, premisas, entradas, salidas, procedimientos y control.

6.1.1. Principios y Requerimientos

La descripción y presentación semántica del CMS UH-WEB está regida por los principios básicos que apoyan su operatividad como herramienta de información. Su diseño es el resultado de la imbricación de conceptos cuyas relaciones en la praxis contribuyen a la obtención de una herramienta para la organización, representación y recuperación de la información en la UH.

Los principios en que se sustenta el modelo son:

- **Consistencia lógica.** En el modelo se establece el proceso a partir una sucesión de pasos de la secuencia planteada y su vinculación en la correspondencia con la lógica de la ejecución de este tipo de estudio.
- **Flexibilidad.** Es un modelo flexible cuyo alcance trasciende las Ciencia de la Información. Permite su aplicación a otras áreas de conocimiento desde el abordaje de la minería de textos, la inteligencia artificial, así como en otros procesos que hoy se estudian en la Ciencia de la Información como vigilancia tecnológica, indización automática, entre otras. Presenta una alta capacidad de actualización y reajuste en los diferentes procesos y procedimientos específicos que plantea.
- **Parsimonia.** Es complejo, tiene un alto nivel de sofisticación. Es coherente con las últimas tendencias y postulados teóricos descritos sobre el tema desde la Ciencia de la Información.
- **Racionalidad.** Es un proyecto racional acorde con la relación costo-beneficio que se requiere para su implementación.
- **Perspectiva o generalidad.** El diseño permite transigir y conformar sistemas de representación de contenidos y permite su utilización como instrumento metodológico para desarrollar otras herramientas de igual nivel de calidad.

Las premisas fundamentales para el desarrollo del modelo son:

- Las bases de datos de las cuales se alimenta el CMS y la Ontología deben estar desarrolladas en MySQL.
- Es necesaria la especificación de bases de conocimiento para el almacenamiento de las anotaciones.

6.1.2. Objetivos

El principal objetivo del CMS que hoy se diseña es facilitar la recuperación y descripción semántica de los contenidos. Este sistema debe permitir la anotación semántica para la realización de búsquedas federadas y búsquedas semánticas en diferentes aplicaciones e integrar un protocolo de cosecha de metadatos para manejar las aplicaciones que lo nutren.

Existen muchas plataformas y CMS al uso en la literatura internacional para el desarrollo y gestión de contenidos. La mayoría de ellas no soluciona el problema de la semántica y la integración de datos. Las aproximaciones más importantes en el terreno de los CMS-Semánticos resuelven los problemas de la semántica usando mecanismos de anotación semántica, que se realizan o bien manuales o de forma automática olvidando las necesidades de intercambio y colaboración.

A continuación se declaran las cualidades y objetivos del CMS UH-WEB.

- **Trabaja estándares del W3C:** El uso de estándares de recuperación de información y de descripción emanados de las normativas del consorcio W3C, ello implica el uso de protocolos de intercambio, descripción y almacenamiento en forma de tripletas de grafo, la utilización de vocabularios de ontologías de alto nivel de especificación y el uso de reglas lógicas e inferenciales.
- **Diseño centrado en el usuario:** El manejo de la web como clave fundamental para que los usuarios se comuniquen con los contenidos, proceso que hasta ahora era mediando por herramientas de elevado costo y gran complejidad.
- **Capacidad para la anotación manual:** Gestión de anotaciones manuales basadas en otros estándares de descripción que aprovechan coherentemente los metadatos de los documentos y de otras entidades no considerados como tal.
- **Anota contenidos semánticos usuales:** La anotación de contenidos que usualmente se anotan en los CMS-Semánticos, entre ellos: páginas web en formato XHTML, imágenes y elementos multimedia.
- **Anota contenidos semánticos inusuales:** La anotación de contenidos que usualmente no se anotan en los CMS-Semánticos, entre ellos: elementos connotativos, denotativos, que se van más allá de la descripción física de un recurso.
- **Consistencia en la anotación:** vela por el seguimiento de las anotaciones semánticas cuando existen modificaciones en la información en los documentos anotados.

- **Normaliza las anotaciones:** Anotaciones homogéneas y normalizadas para cada tipo de recursos en función con las necesidades de las distintas entidades de la UH y la información que estas demandan.
- **Integra anotaciones:** La integración de las anotaciones asociadas al recurso que se describe y almacenadas en una ontología o base de conocimiento.
- **Búsqueda y recupera de información semántica:** Desarrollar tareas de búsqueda y recuperación de información semántica aprovechando las facilidades que brinda el manejo de datos enlazados y en las anotaciones semánticas.
- **Ofrece herramientas para consultas:** Facilidades de las herramientas semánticas para la traducción de las consultas en búsquedas operativas, entendibles por los usuarios.
- **Implementa la UH ontology:** El uso de la ontología para representar los acervos de la organización, clave esencial para el manejo de la interoperabilidad.
- **Originalidad de la propuesta:** la propuesta de CMS es fruto de productos preexistentes combinados de manera original y práctica.

El cumplimiento de los objetivos que se expresaron anteriormente posibilitará que el CMS-UH facilite la organización búsqueda y recuperación de los contenidos asociados a los procesos sustanciales de la Universidad de La Habana, a su vez los datos que se publiquen de estos resultados.

6.1.3. Requisitos de Diseño

En este acápite de la investigación la autora deja plasmado los principios de diseño del CMS Semántico UH-Web, los mismos emanan de las deficiencias encontradas en el proceso de evaluación de la calidad de datos y de las deficiencias constatadas en la valoración por especialistas de las plataformas Joomla y Drupal.

1. Diseño Centrado en el Usuario: El CMS se realizará bajo el principio de la experiencia de usuario, lo que le permitirá a estos el uso de interfaces coherentes y claras para la realización de anotaciones en los documentos de forma colaborativa.

2. Modelado de Ontologías: Como todo CMS semántico las ontologías son la herramienta fundamental para la búsqueda y recuperación de información. En este caso el uso de ontologías se centrará en la estructuración del conocimiento de varios dominios de forma que todos los datos sean tratados coherentemente y además cada ontología deberá ser nombrada etiquetada y definida en el CMS. Los valores de las ontologías estarán acorde con el sistema de anotación que se adopte. El CMS estará preparado para los cambios en las ontologías específicamente para la supresión de clases, la adición de clases y las variaciones en la lógica descriptiva en la ABOX.

3. Soporte de Formatos y Estándares del Web Semántica: Es importante que el SMS maneje los datos en formatos adecuados a las normativas de la W3C, esto facilitará la interoperabilidad con otras aplicaciones y sistemas de información y entre los usuarios, ya que los datos debidamente anotados podrán compartirse en ellos.

4. Automatización: La integración de tecnologías que faciliten la extracción, la publicación, el consumo y el enlazado de datos es vital para el desarrollo del CMS esto permitirá la interoperabilidad semántica y sintáctica con otras aplicaciones, además la anotación semántica debe tener cualidades para el marcado de datos cuya estructura no es describible en DublinCore.

5. Exposición de Datos o OAI-PMH: Es vital que el CMS exponga sus datos en formato OAI-PMH esto lo haría visible para otras aplicaciones externas como el ABCD (Sistema de Gestión de Bibliotecas), el SIGENU, de profesores.

6. Homogeneidad de Datos: El sistema demanda el tratamiento automático de datos, por ello demanda el manejo de automático de mecanismos de limpieza de datos de manera de homogeneizar toda la información y limpiar todas las duplicidades, cuestión que afecta la calidad de los datos que provee.

7. Calidad de Datos: Como todo CMS que se ampara en la representación de datos de organizaciones, el CMS UH-Web se basa en la normativa de calidad de datos y en el resguardo de los mismos.

8. Integración: El sistema está orientado a la integración de componentes en un solo sistema debido a las disponibilidades tecnológicas de la UH.

9. Interfaz de Usuario: La interfaz de usuario debe ser capaz de adaptarse a las cualidades y exigencias de la web semántica en la que los navegadores web no cambian, solo son enriquecidos y adaptados a la filosofía de trabajo.

10. Compatibilidad Multi-navegador: El sistema será compatible con los navegadores más utilizados, para evitar incompatibilidades.

11. Interoperabilidad: El sistema será interoperable sintáctica y semánticamente con la mayoría de las aplicaciones al uso en la Universidad de la Habana.

6.1.4. Arquitectura

Siguiendo los requisitos planteados en el subepígrafe anterior, se define a continuación la arquitectura del CMS UH-WEB. Para ello se presentan dos modelos:

- Modelo lógico del CMS: con la modelación de los procesos que ocurren tras el manejo de la información y los metadatos. (Ver fig.6.1)
- Modelo estructural: con la modelación de las capas técnicas del sistema (Ver Fig.6.5).

6.1.4.1. Modelación lógica del CMS UH-Web

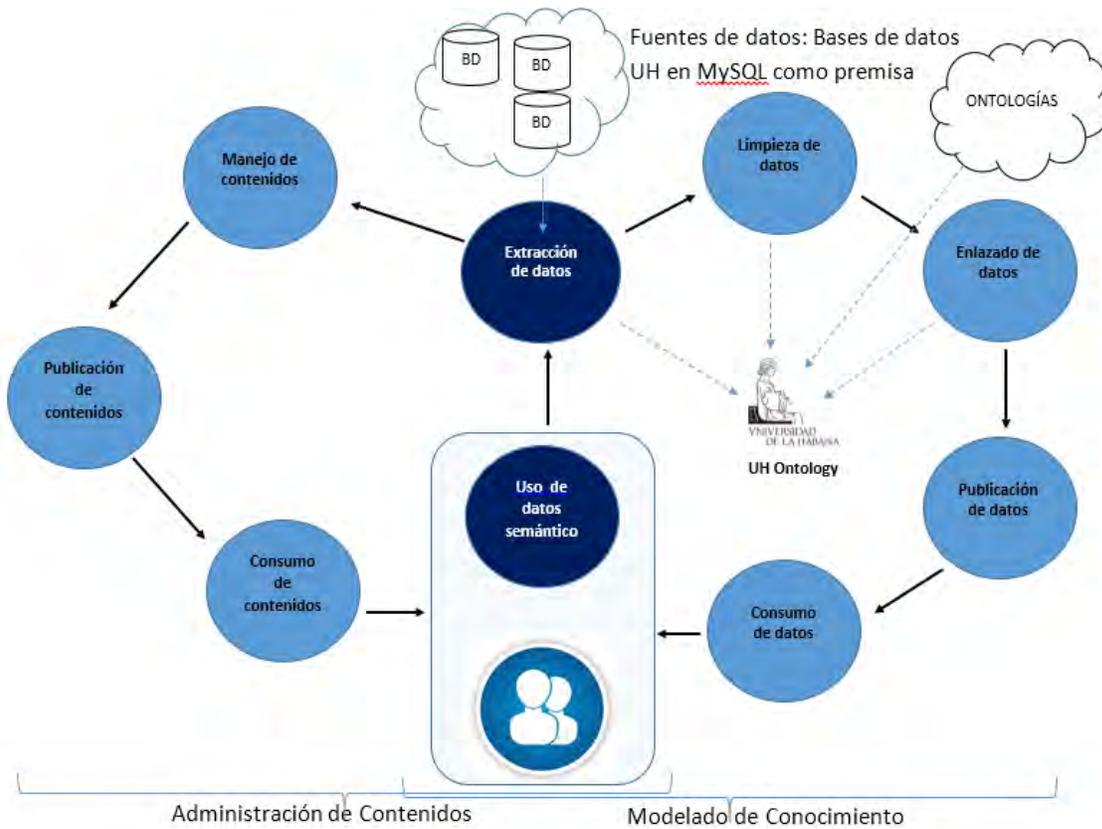


Fig.6.1. Modelo lógico del CMS: UH-WEB (elaboración propia)

El modelo presenta una filosofía cíclica de funcionamiento, ajustado a los principios del Ciclo de Vida de la Información proveniente del campo de las Ciencias de la Información, también conocido como enfoque basado en pipeline, donde la salida de la actividad anterior constituye la entrada a la actividad siguiente. Se divide en dos momentos fundamentales: la **administración de contenidos** y el **modelado de conocimiento**, los que pueden desarrollarse de manera simultánea.

Consta de nueve fases, que se abordan a continuación:

- **Extracción de datos:** el objetivo de esta fase es extraer de las bases de datos de la Universidad de La Habana los datos necesarios para la construcción de una base de conocimientos ontológica.

A nivel tecnológico, para la realización de esta actividad es necesario contar con dos componentes esenciales: (1) recolector OAI-PMH⁷⁶ y (2) motor de base de datos. Por tanto se propone que, a partir de la extracción de los datos de las diferentes plataformas de la Universidad, se declararen reglas de transformación con las que se pueda conseguir que los datos sean convertidos en un formato OAI-PMH para ser almacenados en un servidor web e interrogados mediante un programa en PHP que reciba las peticiones OAI-PMH, aspecto clave para que los datos sean extraídos y convertidos en RDF.

Para realizar este proceso se propone el uso de la herramienta Metharto⁷⁷ (Hidalgo-Delgado, 2015).

Es importante tener en cuenta que los sistemas con que cuenta la UH, por separado, tienen una organización diferente en los datos y en los formatos que los organizan, si bien el DSpace y el Moodle exponen sus datos en OIPMH. Otras fuentes entre las que se encuentran el ABCD (Sistema de Automatización de Bibliotecas y Centros de Documentación) tiene sus datos estructurados en formato ISIS y en XML al contrario de los Sistemas de Gestión (descritos en el capítulo 5) que poseen toda su información almacenada en sistemas de bases de datos relacionales pero no tienen construido el protocolo OAI. Dicha condición

⁷⁶**OAI-PMH**: Open Archives Initiative Protocol for Metadata Harvesting

⁷⁷**MetHarTo** es una herramienta en desarrollo con la finalidad de cubrir las tareas de extracción y preprocesamiento de metadatos diseminados mediante el protocolo OAI-PMH. Es considerado unharvester (aplicación cliente que emite peticiones OAI-PMH, operado por un proveedor de servicios como un medio de recolección de metadatos de los repositorios). Soporta la extracción de metadatos en formato *oai_dc* (DublinCore) aunque, a decir de Hidalgo Delgado (2015), se pretende incorporar otros formatos y estándares tales como: *oai_marc*, *marcxml* y *RFC1807*. El componente es capaz de conectarse a fuentes de datos a través de un proxy. Las opciones de configuración del proxy son proporcionadas mediante el mecanismo de configuración del componente. Recolecta metadatos en lotes y de forma selectiva desde múltiples repositorios. Los metadatos recolectados son almacenados en una base de datos relacional MySQL o PostgreSQL para su posterior uso. A través de **MetHarTo** se genera una base de datos intermedia que luego es utilizada por la herramienta **D2RQ** (BIZER, C. 2004. Andy Seaborne D2RQ-Treating Non-RDF Databases as Virtual RDF Graphs. *Poster at ISWC2004*.) para la generación de tripletas RDF. [HIDALGO-DELGADO, Y. 2015. *Marco de trabajo basado en los datos enlazados para la interoperabilidad semántica en el protocolo OAI-PMH*. Maestría, Universidad de las Ciencias Informáticas.]

obliga homogenizar los datos para que a partir de este protocolo puedan llevarse a formato RDF.

- **Limpieza de datos.** La calidad de los metadatos es fundamental para garantizar la visibilidad, el descubrimiento y la reutilización de los recursos. Esta fase tiene como objetivo limpiar y normalizar los metadatos extraídos en la fase anterior (extracción de datos) para poder definir un modelo ontológico con el fin de compartir y anotar semánticamente, haciéndolos potables tanto para humanos como para los software.

Bajo este fundamento se propone la utilización del sistema Authoris (Leiva Mederos et al., 2013) para homogenizar mediante reglas los nombres de los autores y Hadoop (herramienta que se utiliza para el tratamiento de big data y facilita la homogenización de datos para la construcción de fuentes homogéneas).

Como resultado se crean las condiciones para el modelado de datos, que no es más que la selección de la ontología donde se almacenarán dichos datos. La autora coincide con otros especialistas como (Hidalgo-Delgado, 2015) y (Heath and Bizer, 2011) en la importancia de la reutilización de otras ontologías y vocabularios disponibles, pues una mala selección de las mismas ocasiona errores de sintaxis que luego son arrastrados al proceso de marcado.

Para el diseño del CMS se ha utilizado la ontología VIVO (Steele, 2009) que contiene los siguientes vocabularios ontológicos: bibo, foaf, skos, geonames, vitro, etc.

En este proceso se evalúa la calidad de los datos tomando en cuenta los siguientes indicadores: exactitud, completitud, consistencia, objetividad, interpretabilidad, facilidad de comprensión, consistencia representacional, representación concisa, tiempo de vida, actualidad, volatilidad, veracidad, credibilidad, cantidad apropiada de datos, unicidad, seguridad, utilidad, preservación, comodidad, accesibilidad, objetividad, precisión, relevancia, reputación y unicidad.

- **Enlazado de datos:** El enlazado de datos gravita sobre el desarrollo de dos actividades esencialmente: (1) la construcción del grafo RDF teniendo en cuenta el modelado de los datos establecido en la fase anterior y (2) la generación de

enlaces entre el grafo RDF obtenido y grafos similares. Para ello es imprescindible trabajar en las propiedades que derivan en la sinonimia (como por ejemplo: owl:sameAs⁷⁸ y rdfs:seeAlso⁷⁹).

Para el enlazado se decide utilizar la etiqueta owl:sameAs para unir los diversos recursos de información de los que se provee el CMS.

Los enlaces se realizan mediante Silk⁸⁰ (Isele et al.) que proporciona un lenguaje declarativo de alineación denominado Silk-LSL y facilita el uso de SPARQL Endpoints remotos para realizar el enlazado⁸¹.

Con Slik los desarrolladores pueden especificar qué tipos de enlaces RDF deben ser descubiertas entre las fuentes de datos, así como los elementos que deben cumplir las condiciones de datos con el fin de estar vinculados entre sí. Estas condiciones de enlace pueden combinar diferentes medidas de similitud (ver fig.6.2) y pueden tomar la gráfica en torno a un elemento de datos, utilizando

⁷⁸owl:sameAs: implica que dos URIs (tanto en el grafo origen como en el destino) se refieren al mismo recurso

⁷⁹rdfs:seeAlso: implica que un recurso existente en el grafo de destino proporciona información adicional acerca de un recurso evidente en el grafo de origen.

⁸⁰ SILK- LSL, Es un framework de código abierto para la integración de fuentes de datos heterogéneas. Se basa en el paradigma de Linked Data para el manejo de RDF que proporciona un modelo de datos expresivo para representar información estructurada. En los enlaces RDF se establecen entre entidades de diferentes fuentes de datos.

Es una herramienta que se utiliza para expresar la heurística y decidir si existe una relación semántica entre dos entidades. También puede ser empleado para especificar los parámetros de acceso a las fuentes de datos involucradas, y para configurar las características de almacenamiento en caché, indexación y selección previa. En las condiciones de enlace se pueden utilizar diferentes funciones de agregación para combinar puntuaciones de similitud (VOLZ, J., BIZER, C., GAEDKE, M. & KOBILAROV, G. 2009. Silk-A Link Discovery Framework for the Web of Data. *LDOW*, 538.).

SILK puede ser consultado con mayor detalle en: ISELE, R., JENTZSCH HASSO, A., BIZER, C., VOLZ, J., PETROVSKI, P. & MANNHEIM, U. O. *Silk: The Linked Data Integration Framework* [Online]. Available: <http://silkframework.org/> [Accessed 12 diciembre 2015].

⁸¹Silk, además de las bondades que ofrece, ya ha sido propuesto y utilizado en otros proyectos de universidades cubanas. Tal es el caso de la Universidad de Ciencias Informáticas. Ver proyecto en: HIDALGO-DELGADO, Y. 2015. *Marco de trabajo basado en los datos enlazados para la interoperabilidad semántica en el protocolo OAI-PMH*. Maestría, Universidad de las Ciencias Informáticas.

lenguaje RDF. Silk tiene acceso a las fuentes de datos que deben ser vinculados entre sí a través del protocolo SPARQL. Las especificaciones de enlace pueden ser creadas la interfaz gráfica de usuario o de forma manual en XML (Isele et al.).

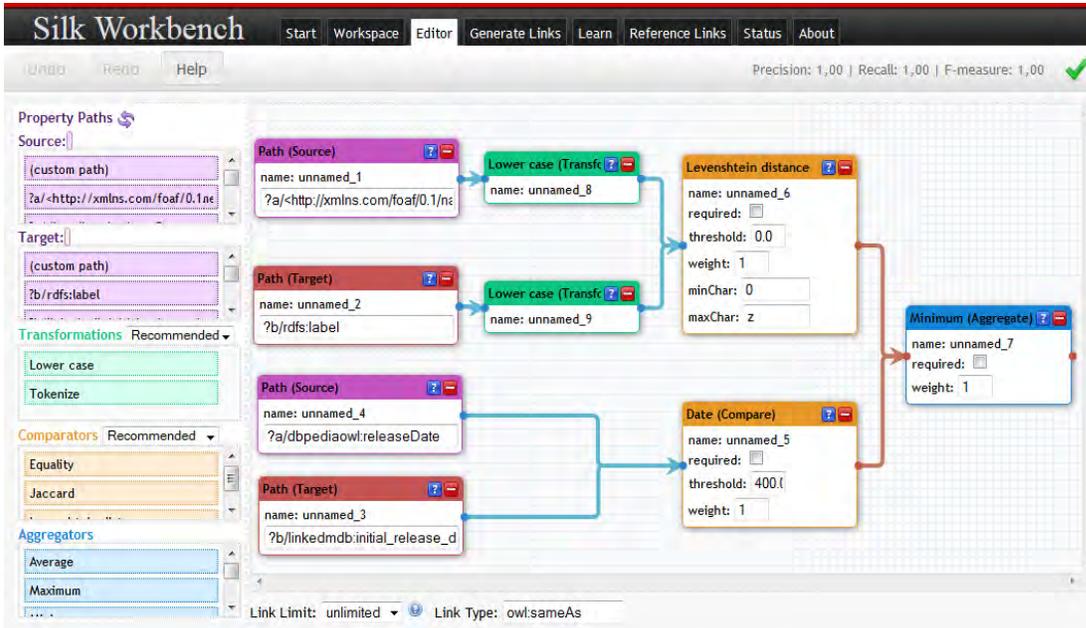


Fig. 6.2. Enlazado de datos desde Silk-LSL (Isele et al.)

Silk además permite al usuario realizar reglas de transformación (Ver fig.6.3) que se pueden utilizar para:

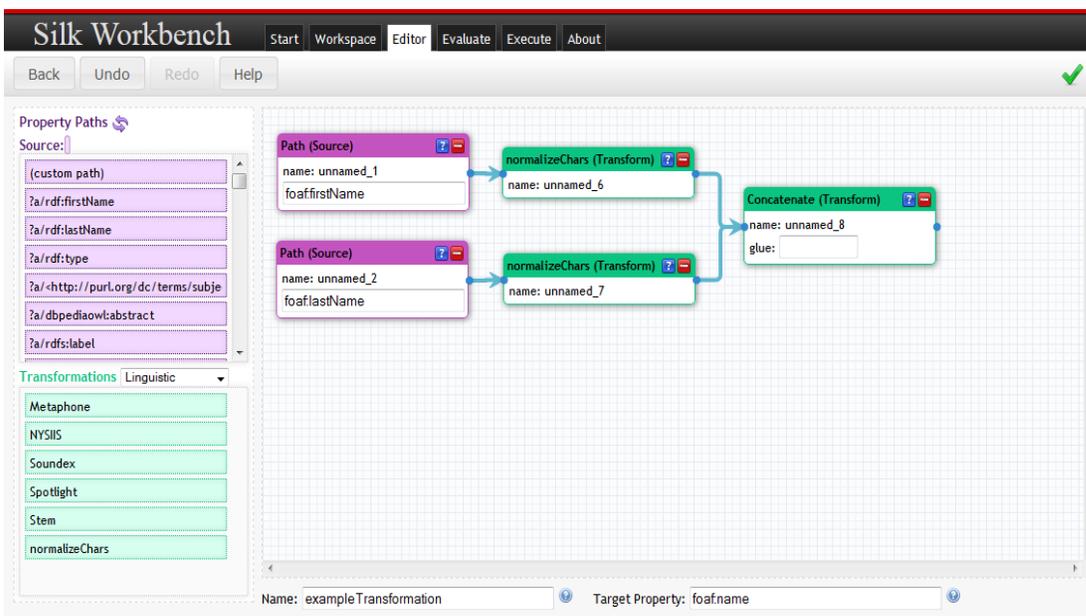


Fig. 6.3. Transformación de datos desde Silk-LSL. (Isele et al.)

- Depuración de los datos
 - Correspondencia entre las diferentes propiedades o la adición de nuevas propiedades con valores generados.
 - La conversión entre diferentes formatos de datos (RDF, CSV o XML).
- **Publicación de datos:** El objetivo de esta fase es la publicación del grafo RDF en la web, garantizando su visibilidad. Por tanto, en esta actividad se trabaja con el modelo ontológico obtenido en la fase de limpieza de datos y con los enlaces RDF ya establecidos en la fase de enlazado de datos.
 - **Consumo de datos:** Una vez trabajados los metadatos, el objetivo de esta fase es el manejo del CMS UH-WEB que se encargará de facilitar la consulta y consumo de datos mediante Navegadores de Datos Enlazados, Motores de Búsqueda de Datos Enlazados y aplicaciones específicas del dominio.

Dentro de esta etapa debe preverse una interface visual que permita al usuario interactuar con la información de la manera más idónea. Para ello desde el CMS UH-WEB se establece la zonificación de los elementos visuales de cada nivel atendiendo a los criterios de jerarquía, usabilidad y estética. Luego se deben establecer definiciones generales de presentación de la información y diseño para acceder a esta, definido ya de antemano el tipo de navegación a utilizar.

- **Uso de datos semánticos:** Esta etapa es un punto trascendental dentro del modelo propuesto, dado que en el confluye el resultado de las etapas anteriores y a la vez se reinicia el ciclo.

Es una etapa que no asume pasos de desarrollo estricto, dada la subjetividad en el proceso de aprendizaje de los individuos que utilizan los datos.

Basados en dicho criterio y partiendo del modelo del ciclo de vida de la información en la web (Caraballo-Pérez and Ramírez-Céspedes, 2007) se asume que en esta etapa ocurren procesos cognitivos diversos a nivel de usuario, donde a partir de la consulta se genera un nuevo conocimiento. (Ver Fig. 6.4)

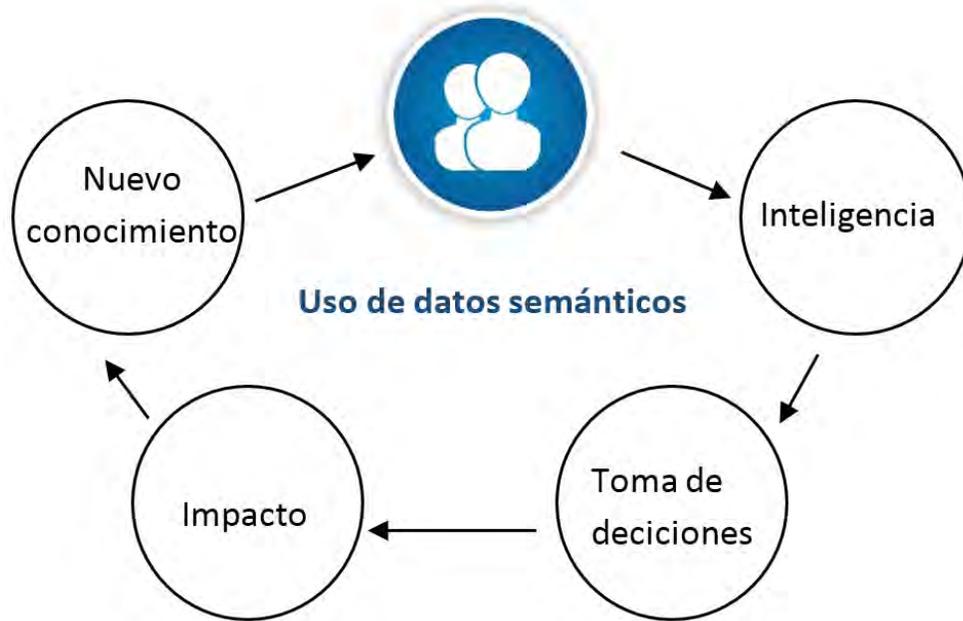


Fig.6.4. Fase de Uso de datos semánticos por parte del usuario (elaboración propia)

Con el esquema anterior (fig.6.4) se intenta modelar los procesos de aprendizaje de cualquier usuario al consultar el sistema y entrar en contacto con la UH ontology, lo que tributa al enriquecimiento mutuo (de proceso cognitivo del individuo y de las relaciones semánticas de la ontología)

Las fases siguientes se desenvuelven de manera similar a lo que ocurre en cualquier sistema gestor de contenidos. Partiendo de la extracción de datos (explicada con anteriormente) hacia el manejo de los contenidos, su publicación, consumo y uso.

A diferencia de la vertiente anteriormente explicada (ver fig.6.1), en las fases que se abordan a continuación se parte del principio del manejo de contenidos y no de los metadatos. La gestión de contenidos está orientada al manejo de objetos que actúan como componentes de documentos digitales. Los objetos son tratados mediante un conjunto de procesos estructurados con la finalidad de producir publicaciones digitales basadas en la metáfora del documento.

- **Manejo de contenidos:** En esta etapa se reciben los documentos y sus componentes, asumidos (todos) como objetos digitales

Debe dar soporte a los procesos de creación de contenidos, soporte a flujos de trabajo, sindicación e integración de fuentes externas. Además, debe ofrecer

soporte a procesos de conversión entre formatos diversos, y a la agregación de contenidos de fuentes en estructuras específicas.

Etapas donde se realizan los procesos de gestión y control de los repositorios de información, los grupos de usuarios, etc. Se encarga de definir y controlar los flujos de trabajo que son utilizados (en otras etapas del modelo), y de la definición de parámetros para el funcionamiento del sistema.

- **Publicación de contenidos:** En el proceso de publicación de contenidos digitales se define el diseño de un marco de integración para el conjunto de objetos cuyo resultado final es la publicación del documento digital.

En esta etapa se lleva a cabo la producción final de publicaciones o productos de información digital, de manera automática o casi automática. Deberá ofrecer posibilidades de personalización para usuarios y la posibilidad de producir para diferentes tipos de plataformas y/o clientes.

- **Consumo de contenidos:** Se desarrollan los mismos procesos que en la fase de consumo de datos, pero a diferencia de esta, se gestionan los documentos digitales y sus diferentes componentes con visión de gestión de objetos. Esta fase antecede a la ya explicada fase de consumo de datos semánticos.

6.1.4.2. Modelación estructural del CMS UH-Web

El modelo estructural se basa en la filosofía de construcción de N capas y establece cuatro capas técnicas (ver fig.6.5). Es un esquema simple cuya lectura se produce desde la base hacia las capas superiores, característica similar a las establecidas en otros modelos estudiados.

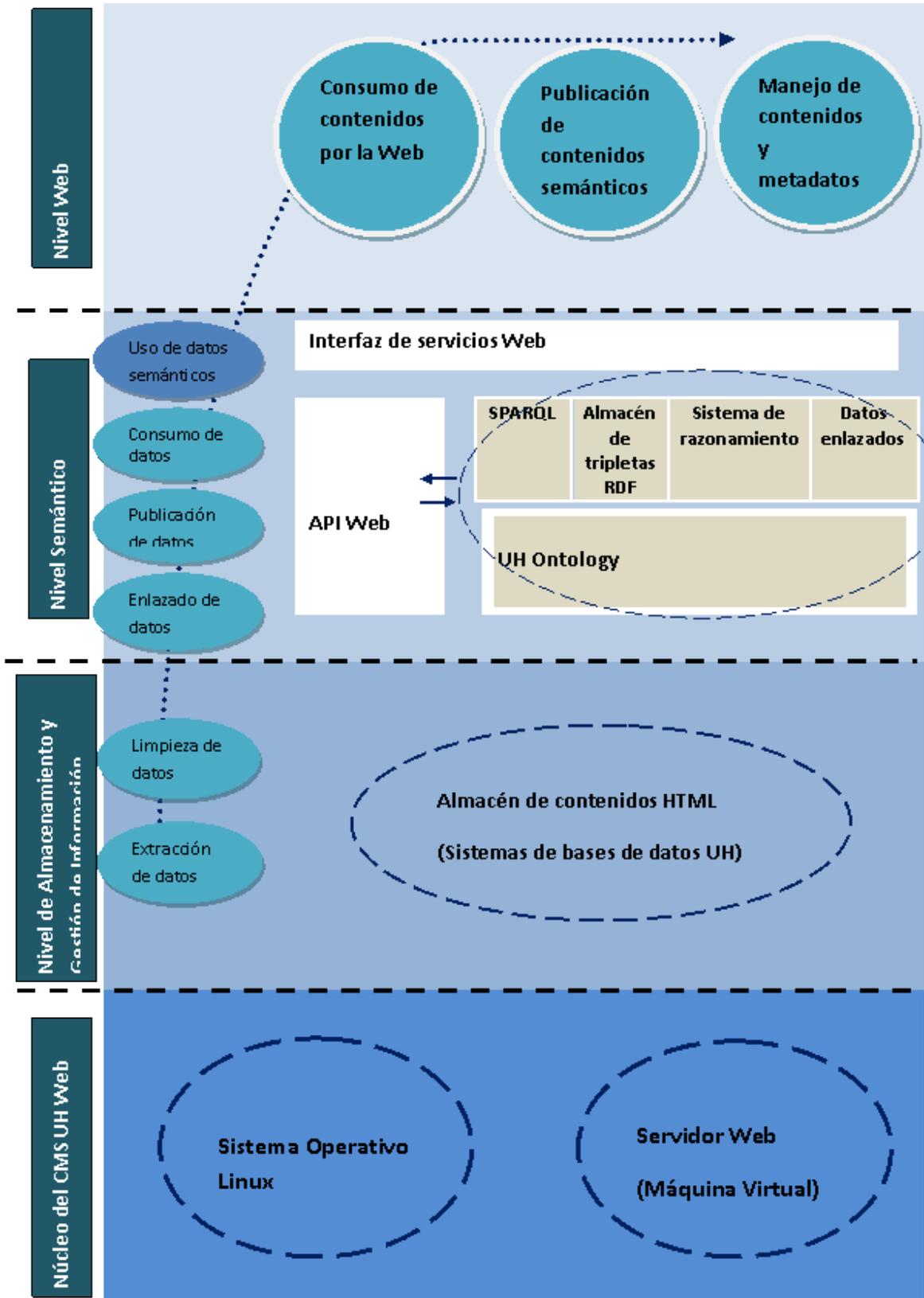


Fig. 6.5. Modelo estructural del CMS: UH-WEB (elaboración propia)

- **Nivel de Núcleo del CMS:** aporta la infraestructura del software y de comunicaciones necesaria para el funcionamiento del resto de los componentes del CMS. Se establece como sistema operativo se propone Linux para garantizar que se desarrolle en el marco del software libre; y una máquina virtual como servidor Web.
- **Nivel de Almacenamiento y Gestión de Información:** en este nivel se desarrollan dos de las etapas establecidas en el modelo lógico del CMS: la extracción y limpieza de datos (ver figs. 6.1 y 6.5). Se establece el almacenamiento de los contenidos HTML de los datos y contenidos almacenados en las diferentes bases de datos (en MySQL) de la UH.
- **Nivel Semántico:** Capa que se establece para extraer el conocimiento de los contenidos. En este nivel se desarrollan cuatro de las etapas establecidas en el modelo lógico del CMS: enlazado, publicación y consumo de datos; y el uso de datos semánticos desde los servicios de aplicación del servidor y a través de la interfaz Web (ver figs. 6.1 y 6.5). En él confluyen los servicios de contenidos encargados de trabajar con la información dirigida a los usuarios de la UH; con los servicios semánticos cometidos al enriquecimiento de dicha información con los metadatos dirigidos a las computadoras.

Se establecen consultas a UH ontology mediante SPARQL, almacenamiento de tripletas RDF, sistema de razonamiento y datos enlazados, los que unidos a los datos UH ontology, se mantendrán en constante interacción con API Web para la gestión de los datos semánticos.

- **Nivel Web:** en este nivel se realiza toda la interacción con el CMS UH WEB. Se desarrollan tres de las etapas del modelo lógico del CMS: consumo de contenidos, publicación de contenidos semánticos y manejo de contenidos y metadatos (sintácticos y semánticos). Es la capa que permitirá la presentación del conocimiento al usuario final y apoyar la interacción directa con este conocimiento.

A diferencia de otros trabajos discutidos en el Capítulo 3 dedicado al tema, la propuesta de la autora no solo establece capas técnicas del CMS con niveles para el tratamiento de la información. Se propone además, un modelo lógico con

la representación de procesos ocurren de manera simultánea en función de las necesidades informativas del usuario.

6.1.5. Ontología y lenguajes ontológicos

Todo CMS dispone de una fuente que va a distinguir el desarrollo de los procesos de gestión. Para el caso de los CMS semánticos son las ontologías. Por dicha razón para el CMS personalizado UH-WEB se ha diseñado UH ontology. Sus características, requerimientos y filosofía de construcción se explican a continuación.

6.1.5.1. UH ontology

La UH ontology es extensa y compleja, característica coherente con las particularidades de la UH, entidad para la cual se ha diseñado dicha ontología.

A partir de los criterios expuestos en el capítulo 2, podría definirse como una ontología de dominio, mixta (dado que combina aspectos gramáticos, semánticos y sintácticos destinados a la utilización de los trabajadores y estudiantes de la universidad, pero también a las máquinas para lograr una adecuada gestión semántica de los contenidos) y lógica.

UH Ontology ha sido pensada como herramienta de referencia en la construcción de la base de conocimiento, contribuir y aportar consistencia, fiabilidad, para combatir el exceso de ambigüedad en la recuperación de información en los sitios Web de la UH. La ontología va a determinar la estructura de la información; la arquitectura de los objetos, sus dependencias y su representación gráfica; incluyendo el almacenamiento de información en una base de datos semántica en forma persistente.

La ontología UH ontology podrá responder:

- PREGUNTAS SOBRE EL DOMINIO: permitirá modelar (representar) los conceptos manejados en la UH con respecto a personas, documentos y espacios geográficos, de manera que sean utilizados por el CMS y entendibles por la máquinas.
- PREGUNTAS SOBRE EL USO: esta ontología facilitará la búsqueda de recursos de información dentro de la UH.

- USO: los usuarios potenciales de la ontología son los estudiantes, profesores, personal administrativo y directivos de la UH.
- MANTENIMIENTO: Los administradores de la ontología serán las personas encargada del mantenimiento de la misma. Se prevee que potencialmente sean especialistas de la Dirección de Informatización, la Dirección de Información y personal colaborador de la Facultad de Matemática y Computación de la propia UH.

Las funciones fundamentales de la ontología que se propone serán:

- Normalizar los atributos de los metadatos aplicables a los documentos.
- Crear una red de relaciones que aporte especificación y fiabilidad.
- Compartir conocimiento en pos de hacer explícito diferentes criterios, así como la integración de diferentes perspectivas de los usuarios.
- Posibilitar el trabajo cooperativo al funcionar como soporte común de conocimiento entre las distintas áreas y estructuras de la UH.
- Tratamiento ponderado del conocimiento para recuperar información de forma automatizada.
- La reutilización del conocimiento existente en nuevos sistemas.
- La interoperabilidad entre sistemas distintos sistemas existentes en la UH.
- Establecer modelos normativos que permitan la creación de la semántica de un sistema y un modelo para poder extenderlo y transformarlo entre diferentes contextos.

La ontología cuenta con 191 clases, 92 propiedades definidas, 60 tipos de datos, entre otras características apreciables en la figuras siguientes. (ver Fig. 6.6 a 6.11).

Ontology metrics:	
Metrics	
Axiom	1609
Logical axiom count	437
Class count	166
Object property count	93
Data property count	59
Individual count	80
DL expressivity	ALCHI(D)

Fig. 6.6. UH ontology: Conteo de las estructuras principales

Class axioms	
SubClassOf axioms count	171
EquivalentClasses axioms count	2
DisjointClasses axioms count	0
GCI count	0
Hidden GCI Count	1

Fig. 6.7. UH ontology: Axiomas

Object property axioms	
SubObjectPropertyOf axioms count	20
EquivalentObjectProperties axioms count	0
InverseObjectProperties axioms count	2
DisjointObjectProperties axioms count	0
FunctionalObjectProperty axioms count	0
InverseFunctionalObjectProperty axioms count	0
TransitiveObjectProperty axioms count	0
SymmetricObjectProperty axioms count	0
AsymmetricObjectProperty axioms count	0
ReflexiveObjectProperty axioms count	0
IrreflexiveObjectProperty axioms count	0
ObjectPropertyDomain axioms count	33
ObjectPropertyRange axioms count	26
SubPropertyChainOf axioms count	0

Fig.6.8. UH ontology: Objetos

Data property axioms	
SubDataPropertyOf axioms count	26
EquivalentDataProperties axioms count	9
DisjointDataProperties axioms count	0
FunctionalDataProperty axioms count	0
DataPropertyDomain axioms count	35
DataPropertyRange axioms count	35

Fig. 6.9. UH ontology: Datos

Individual axioms	
ClassAssertion axioms count	42
ObjectPropertyAssertion axioms count	18
DataPropertyAssertion axioms count	18
NegativeObjectPropertyAssertion axioms count	0
NegativeDataPropertyAssertion axioms count	0
SameIndividual axioms count	0
DifferentIndividuals axioms count	0

Fig. 6.10. UH ontology: Personas

Annotation axioms	
AnnotationAssertion axioms count	756
AnnotationPropertyDomain axioms count	0
AnnotationPropertyRangeOf axioms count	0

Fig. 6.11. UH ontology: Anotaciones

Los nombres de las clases y subclases se trabajaron desde el idioma inglés asignando label en español, para facilitar la coherencia entre los vocabularios manejados y la estructura general de la ontología.

6.1.5.1.1. Reutilización de ontologías

Se toma como base para el desarrollo de UH ontology la ontología VIVO. Dentro de esta última se crearon clases en función de las particularidades de la UH, como por ejemplo Learning, tomada de la IEEE.

En el proyecto UH-Ontology se han reutilizado:

- SKOS (Solomou and Papatheodorou, 2010), con funciones de tesaurus.
- FOAF (Brickley and Miller, 2010), que sirve de herramienta de control de autoridades.
- BIBO (Dimić Surla et al., 2012) Se utiliza para describir objetos bibliográficos en la web semántica en RDF. Esta ontología se puede utilizar como una ontología citación, como una clasificación ontología documento, o simplemente como una forma de describir cualquier tipo de documento en RDF.
- VIVO (Jcorson-Rikert, 2011), es una herramienta para la representación de información sobre la investigación y los investigadores (sus trabajos académicos, líneas de investigación, y las relaciones de organización).
- Geontology (Larin-Fonseca and Garea-Llano, 2011) es un enfoque de integración semántica de datos geoespaciales, utilizados en la gestión de datos geográficos. Adaptación de Geopolitical ontology de la FAO (<http://aims.fao.org/aos/geopolitical.owl>)

6.1.5.1.2. Jerarquía de Clases

Las clases principales establecidas para UH ontology son: Agent, Collection, Document, Resource, Concept, Credential, EducationalTraining, Event, Project, Area, Seq y Software. En la figura 6.12, se representan dichas clases, en un primer nivel de subordinación desde el núcleo (Thing).



Fig.6.12. UH ontology: Clases principales. Visualización desde Protegé.

Las clases que componen la ontología serán abordadas a continuación.

→ **Agent:**

La clase Agent se utiliza para describir cualquier "agente" en relación con los artículos bibliográficos. Dichos agentes pueden ser personas, organizaciones o grupos de cualquier tipo. Por tanto, contiene dos subclases (ver Fig. 6.13): **Organization** (dedicada a la identificación de las diferentes estructuras y sus subordinaciones. Ver Fig.6.15) y **Person** (dedicada a establecer los roles o funciones que puedan tener las personas en su relación con la UH y los diferentes recursos de información gestionados. Ver Fig.6.16)

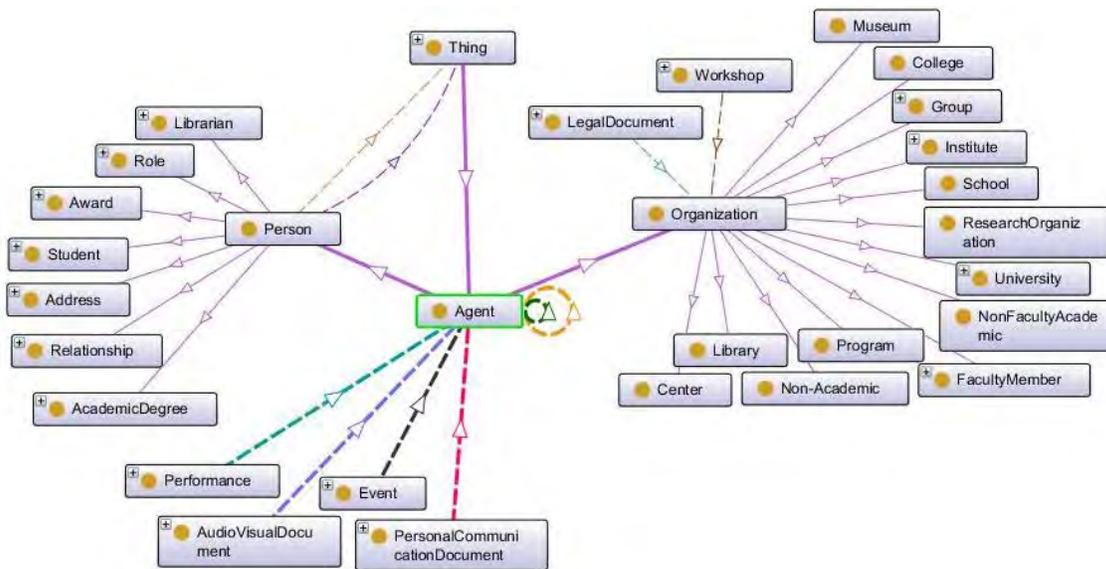


Fig.6.13. UH ontology: Clase Agente y su relación con otras clases y subclases de la ontología (1). Visualización desde Protégé

Además se establecen relaciones entre Agent como rango y otras clases y subclases como dominio (Ver fig.6.13 y 6.14): con la clase Event (relación organizer) y su subclase Performance (relación performer) y con AudioVisualDocument (relación director) y PersonalCommunicationDocument (relación recipient), ambas subordinadas a la clase Document. Se establecen relaciones dentro de la misma clase Agent de tipo interviewee e interviewer.

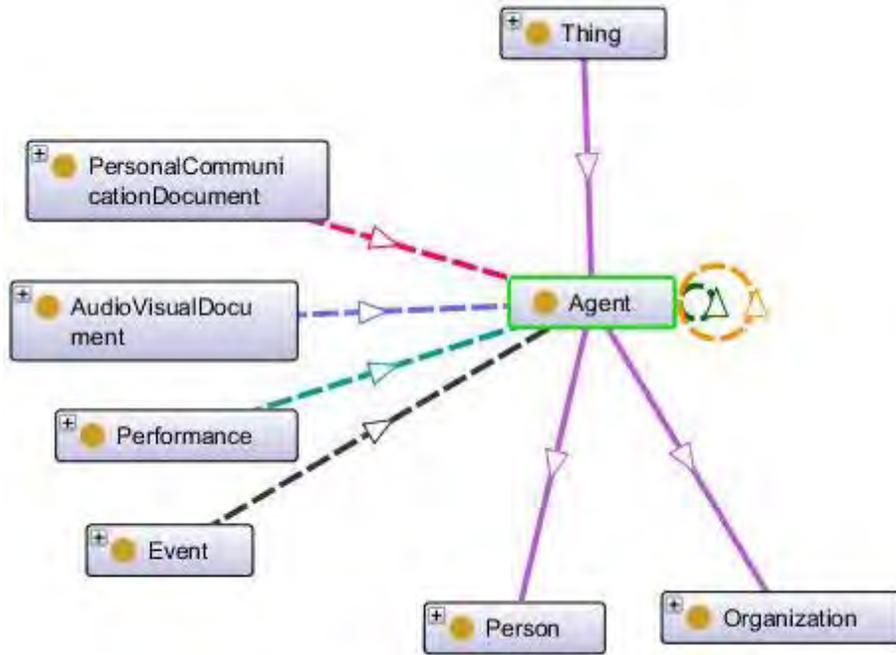


Fig. 6.14. UH ontology: Clase Agente y su relación con otras clases y subclases de la ontología (2). Visualización desde Protegé.

Dentro de *Agente* se encuentra la subclase *Organization* la cual permite presentar en su jerarquía de clases (Ver Fig.6.15) la complejidad organizacional de la UH, dada la variedad de estructuras representadas en la ontología.

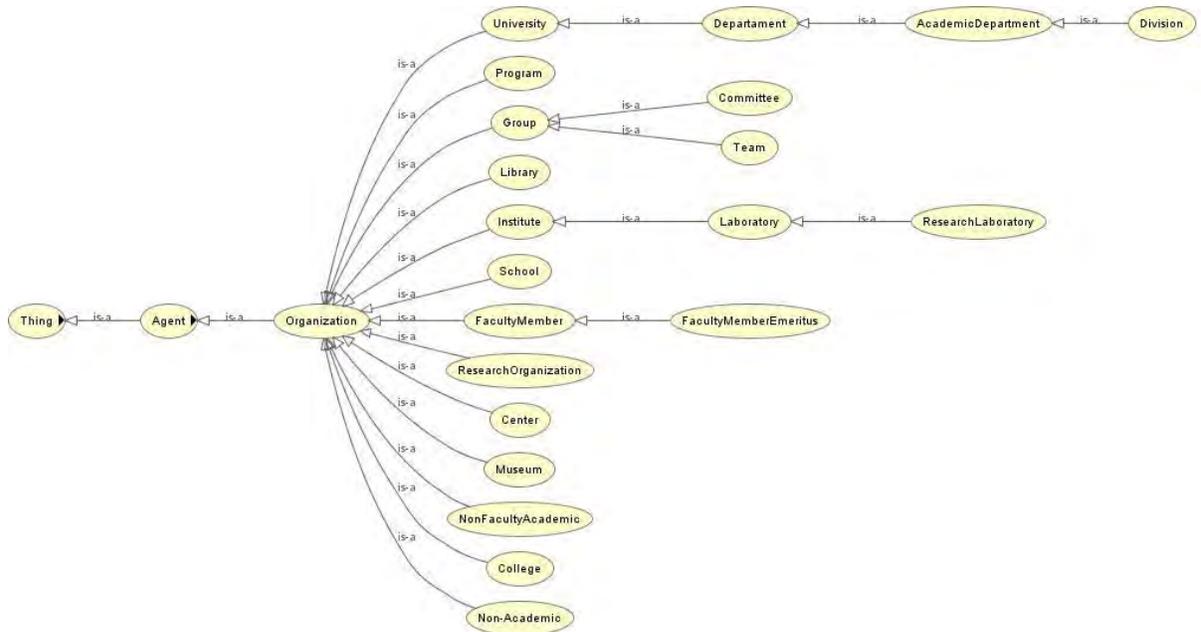


Fig. 6.15. UH ontology: Relaciones de subordinación de la subclase Organización

Cómo puede apreciarse en las figuras 6.13, 6.15 y 6.16 se establecen relaciones de jerarquía de clases. De la subclase *Organization* se derivan: *University*, *Program*, *Group*, *Library*, *Institute*, *School*, *FacultyMember*, *ResearchOrganization*, *Center*, *Museum*, *NonFacultyAcademic*, *College*, *NonAcademic*. Todas ellas establecidas para permitir una coherencia entre la ontología y las diversas estructuras de la UH.

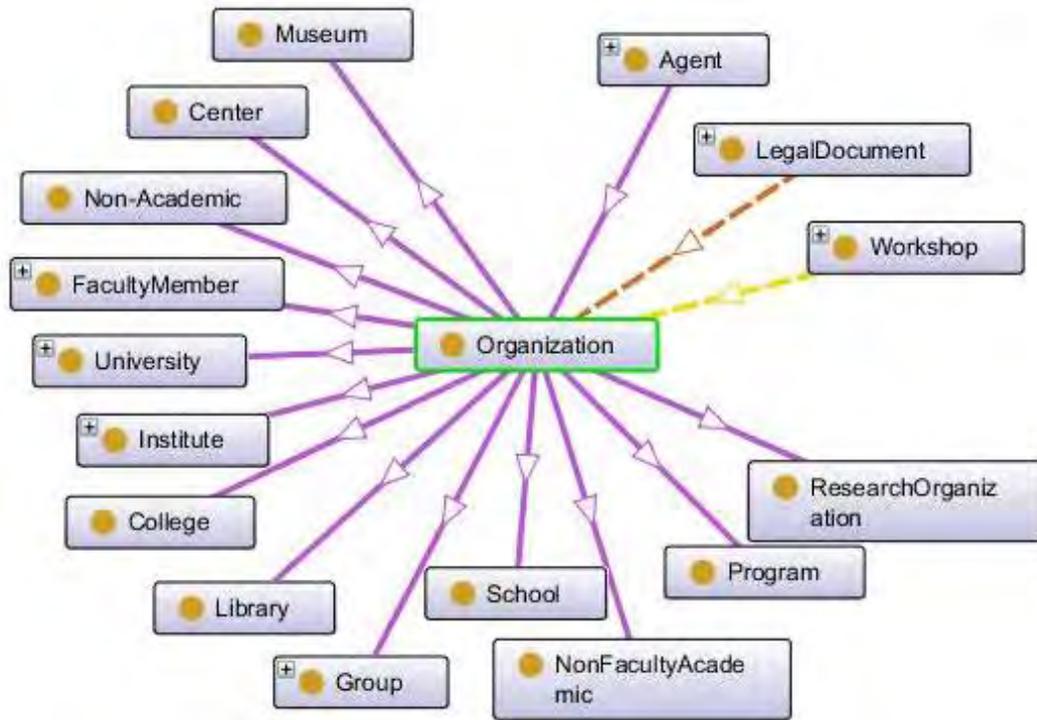


Fig. 6.16. UH ontology: subclase Organización y relación con otras clases y subclases. Visualización desde Protegé.

Además de las relaciones de subordinación (ver en la fig. 6.16 los enlaces mediante las líneas continuas), la subclase *Organization* establece otros vínculos (como dominio) con las subclases *LegalDocument* (court) de la clase *Document* y con *Workshop* (assignedby) de la clase *Event*.

La subclase *Person*, contenida al igual que *Organization* en la clase *Agent*, presenta en su jerarquía de clases (Ver Fig.6.17) la estructura y escalabilidad necesaria para describir al usuario ubicado en la UH desde su posición de trabajador y/o estudiante.

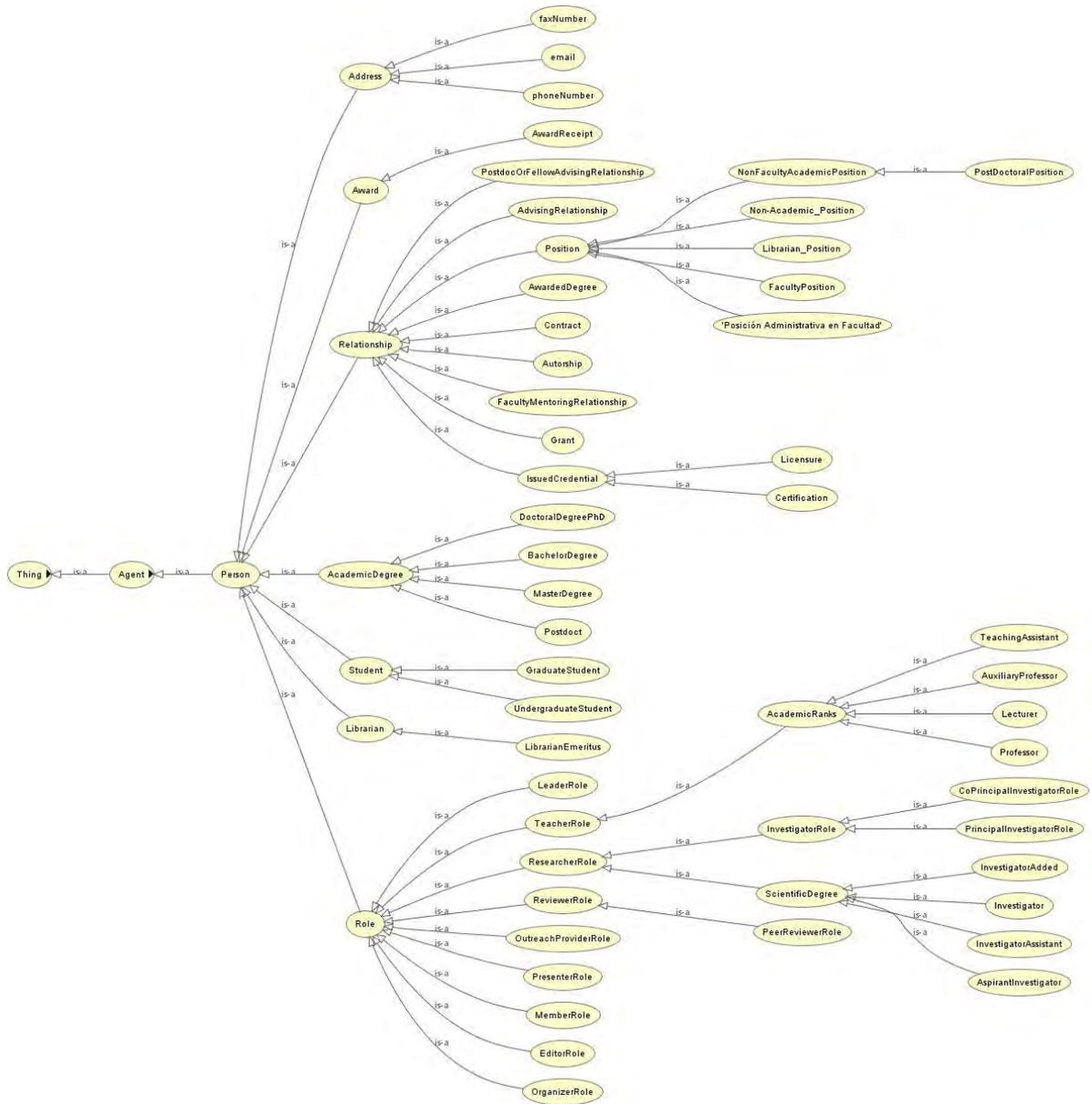


Fig.6.17. UH ontology: Relaciones de la subclase Person

Además de las relaciones de subordinación de la subclase Person (enlaces mediante las líneas continuas), se establecen otros vínculos como dominio con Thing en calidad de rango, para las relaciones IsReferencedBy y hasPrincipallInvestigatorRole.

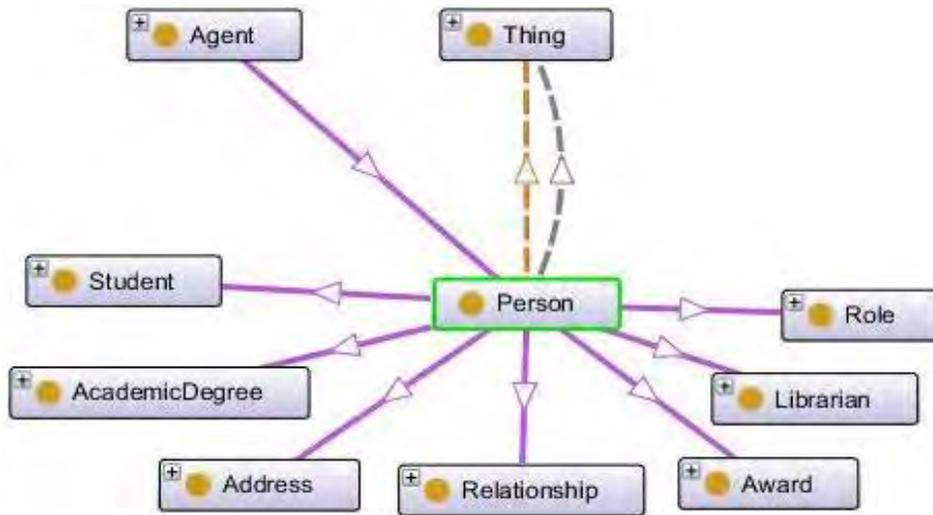


Fig. 6.18. UH ontology: subclase Person y su relación con otras clases y subclases. Visualización desde Protegé

→ **Collection, Document, Resource, InformationResource:**

Se definen a sí mismas de manera independiente pero siempre bajo el principio de declarar su similitud conceptual (Ver fig.6.19).

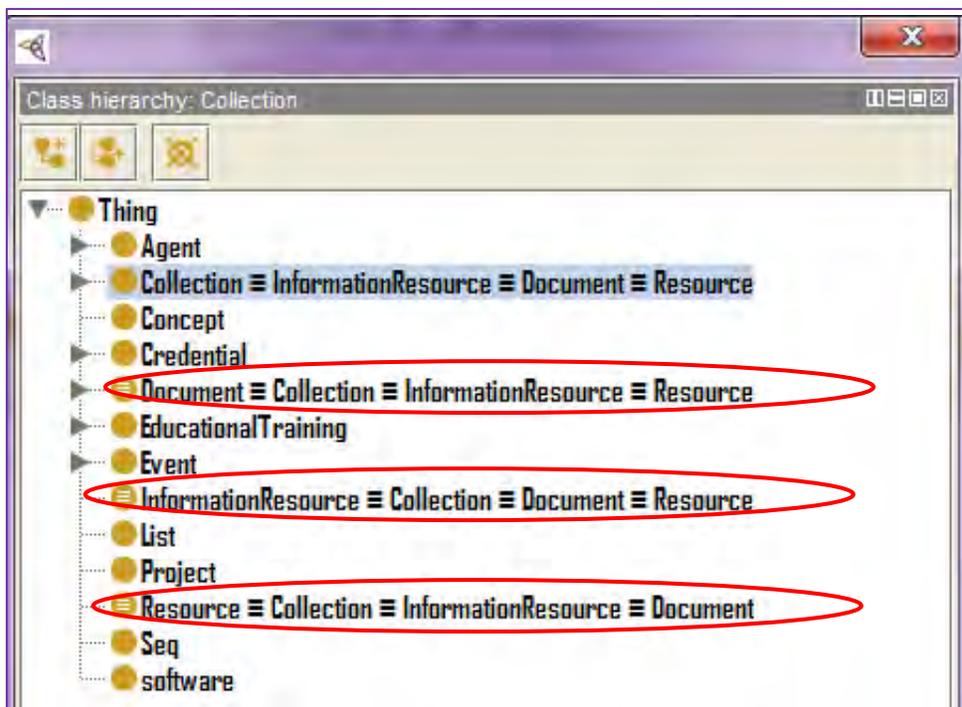


Fig. 6.19. UH ontology: Collection, InformationResource, Document y Resource. Visualización desde Protegé

Dadas las implicaciones de los términos en el ámbito de la Bibliotecología y la Documentación en el reconocimiento y tratamiento de la información, se realiza la declaración de subclases para las clases de Collection (ver fig.6.20) y Document. Los nexos conceptuales establecidos para Document, Resource, Collection e InformationResource son claramente perceptibles a través de la figura 6.22.

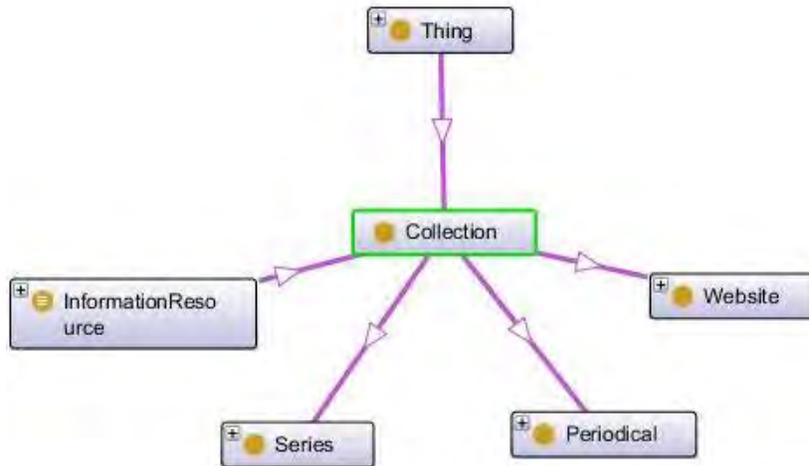


Fig. 6.20. UH ontology: clase Collection. Visualización desde Protegé

Para la clase Collection se establecen relaciones de jerarquía pero para las subclases definidas en el se establecen otras relaciones (hasPart) representadas en la Fig.6.21.

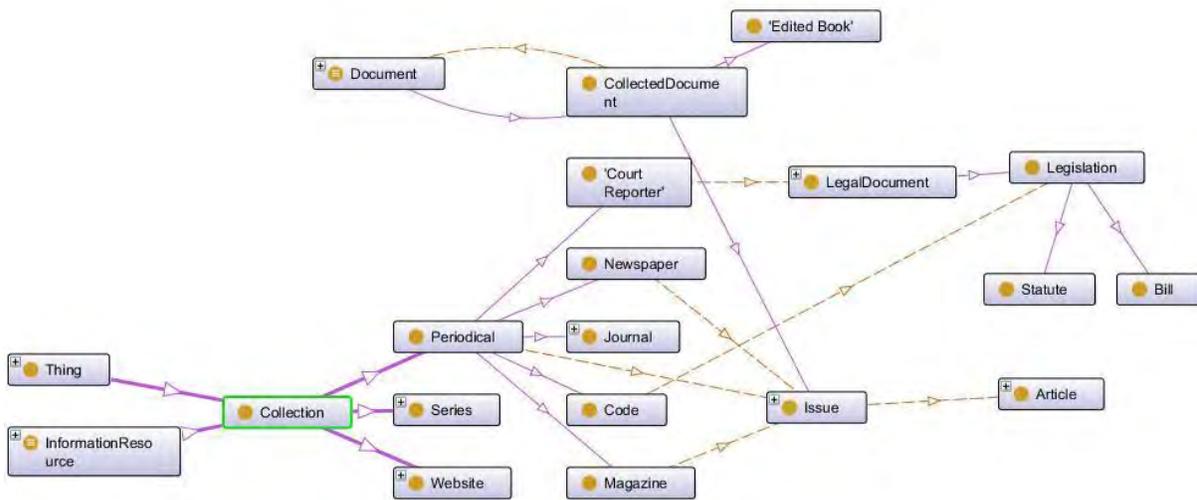


Fig. 6.21. UH ontology: clase Collection y relaciones entre sus subclases. Visualización desde Protegé

No obstante como se declaró anteriormente los terminos colección, recurso de información, documento y recurso estrán estrechamente relacionados y así queda reflejado en UH ontology (ver fig. 6.22, 6.25 y 6.26)

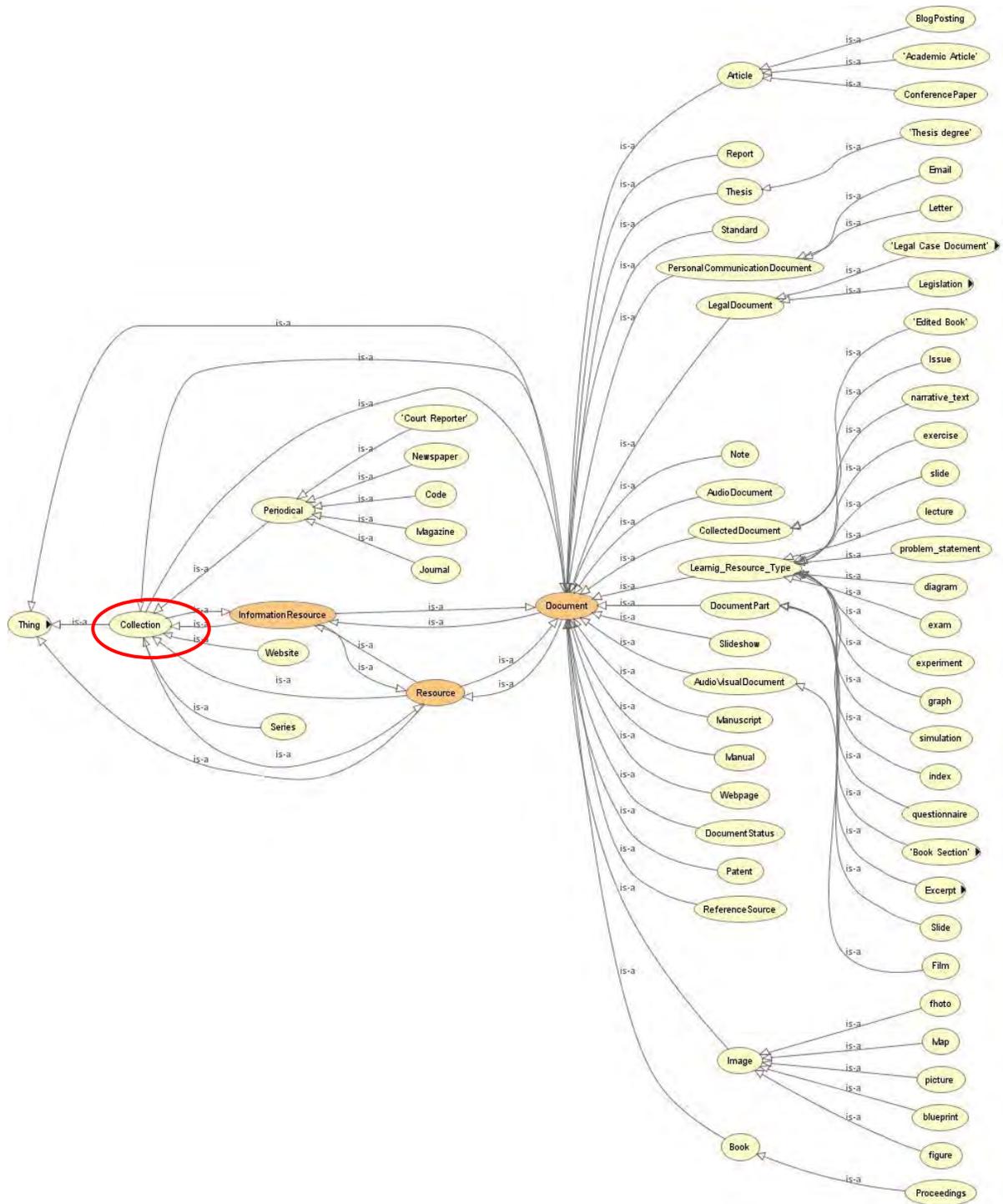


Fig. 6.22. UH ontology: Relaciones de la clase Collection

La clase *Document* presenta varias subclases: *Article*, *AudioDocument*, *AudioVisualDocument*, *Book*, *CollectedDocument*, *DocumentPart*,

DocumentStatus, *Image*, *Learnig_Resource_Type*, *LegalDocument*, *Manual*, *Manuscript*, *Note* y *Patent*. Cada una a su interior posee otras subclases (ver Fig.6.23)

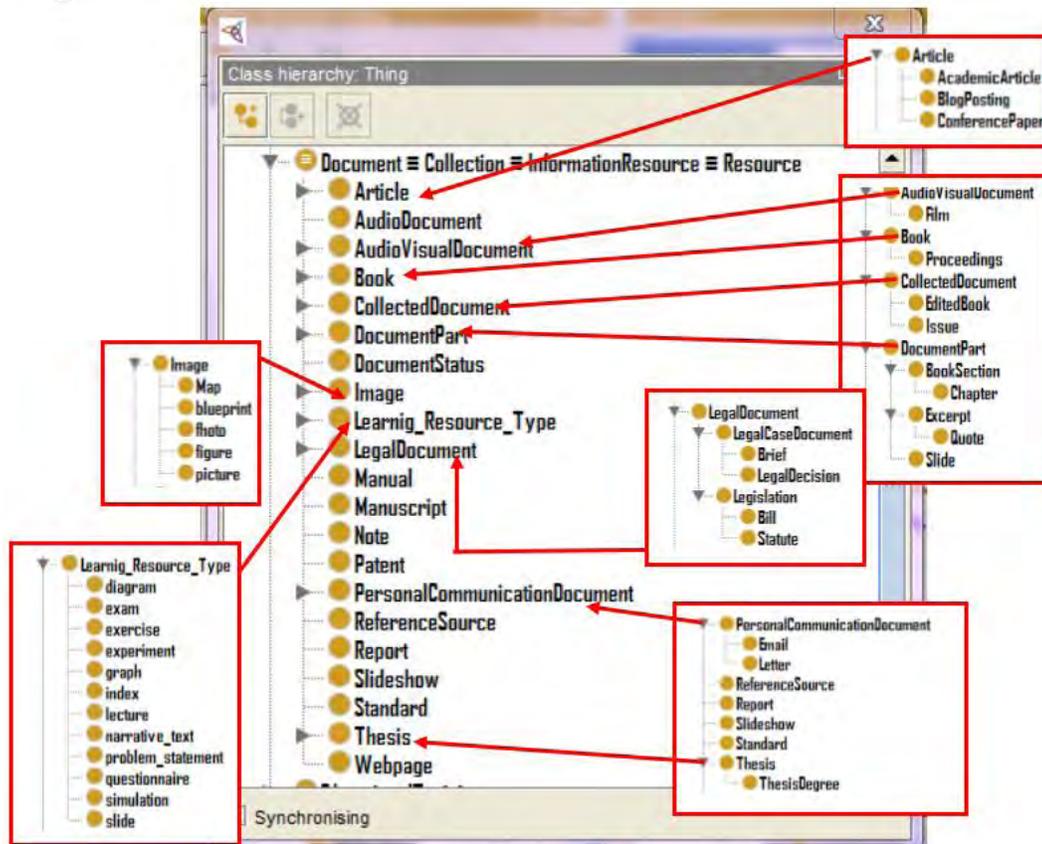


Fig. 6.23. UH ontology: Subclases definidas dentro de Document. Visualización desde Protegé y adaptación propia

Además de las relaciones de subordinación de la clase Document (ver en la fig. 6.24), se establecen otros vínculos en calidad de rango y dominio sobre ella misma para las relaciones de *citedBy*, *reproducedIn*, *translation on* y *cites* (representadas en la fig.6.24 en forma circular), con Thing en calidad de dominio para las relaciones 'presented at' e *IsReferencedBy* y en calidad de rango para las relaciones: *isVersionOf*, 'list of authors', 'list of contributors', 'presented at' y 'list of editors'. Se establece relaciones de tipo 'review of' y 'transcript of' con la clase Resource (esta última en calidad de rango). *CollectedDocument* y *Series* se definen como *hasPart* (para toda subclase de documento).

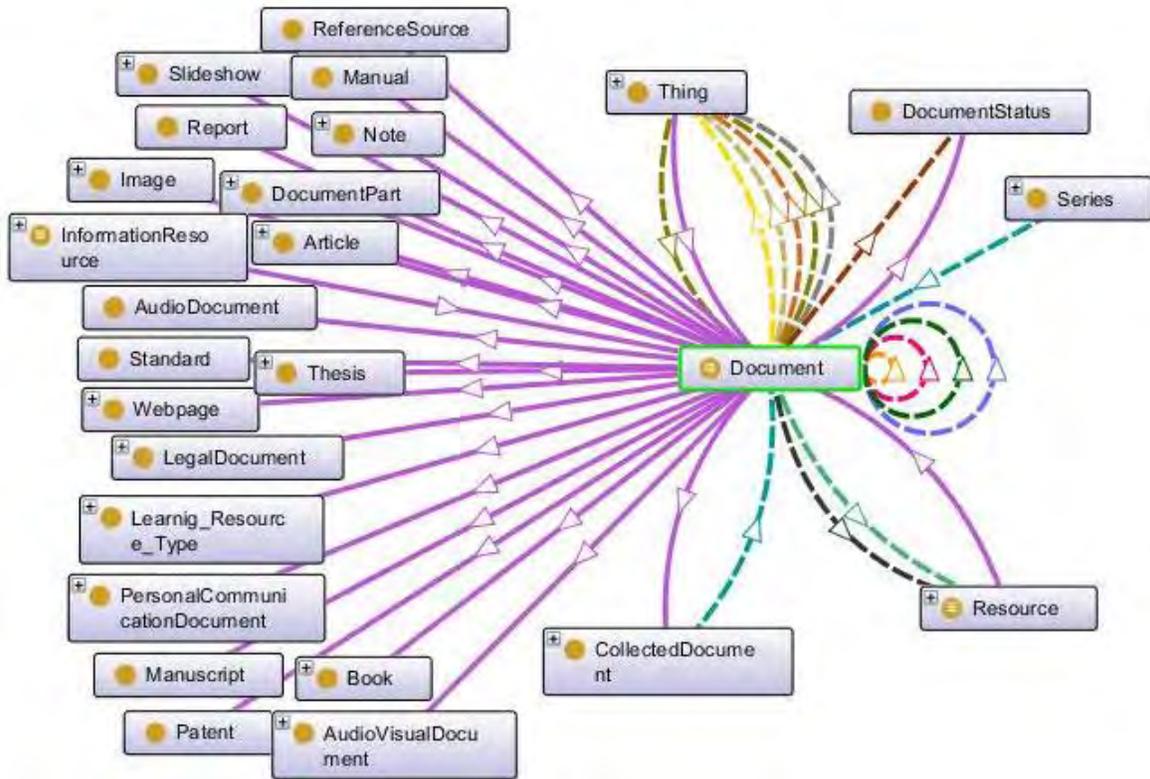


Fig.6.24. UH ontology: Relaciones de la clase Document. Visualización desde Protegé

Los nexos conceptuales establecidos para Document, Resource, Collection y InformationResource son claramente perceptibles a través de las figuras 6.25 y 6.26.

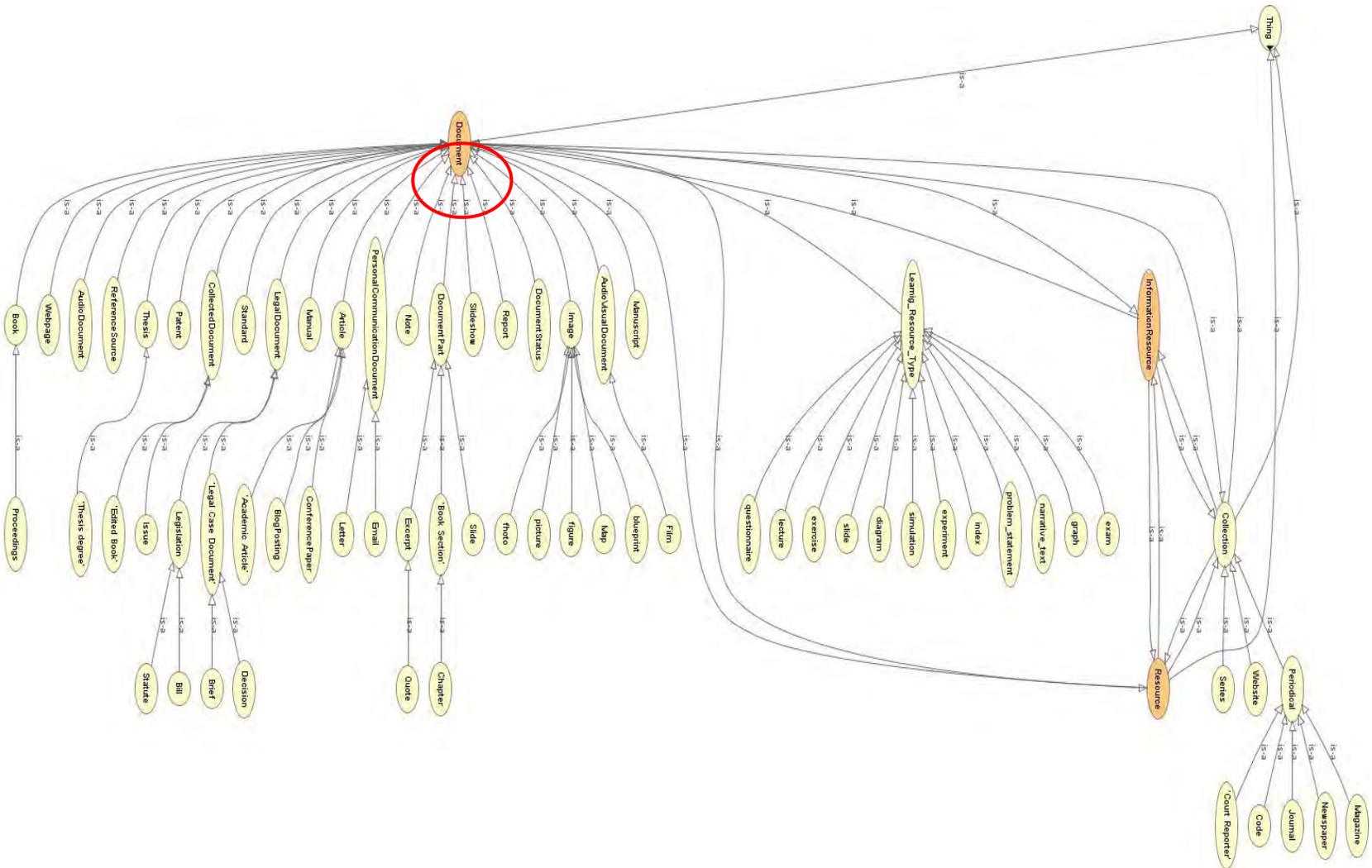


Fig. 6.25. UH ontology: Relaciones de la clase Document

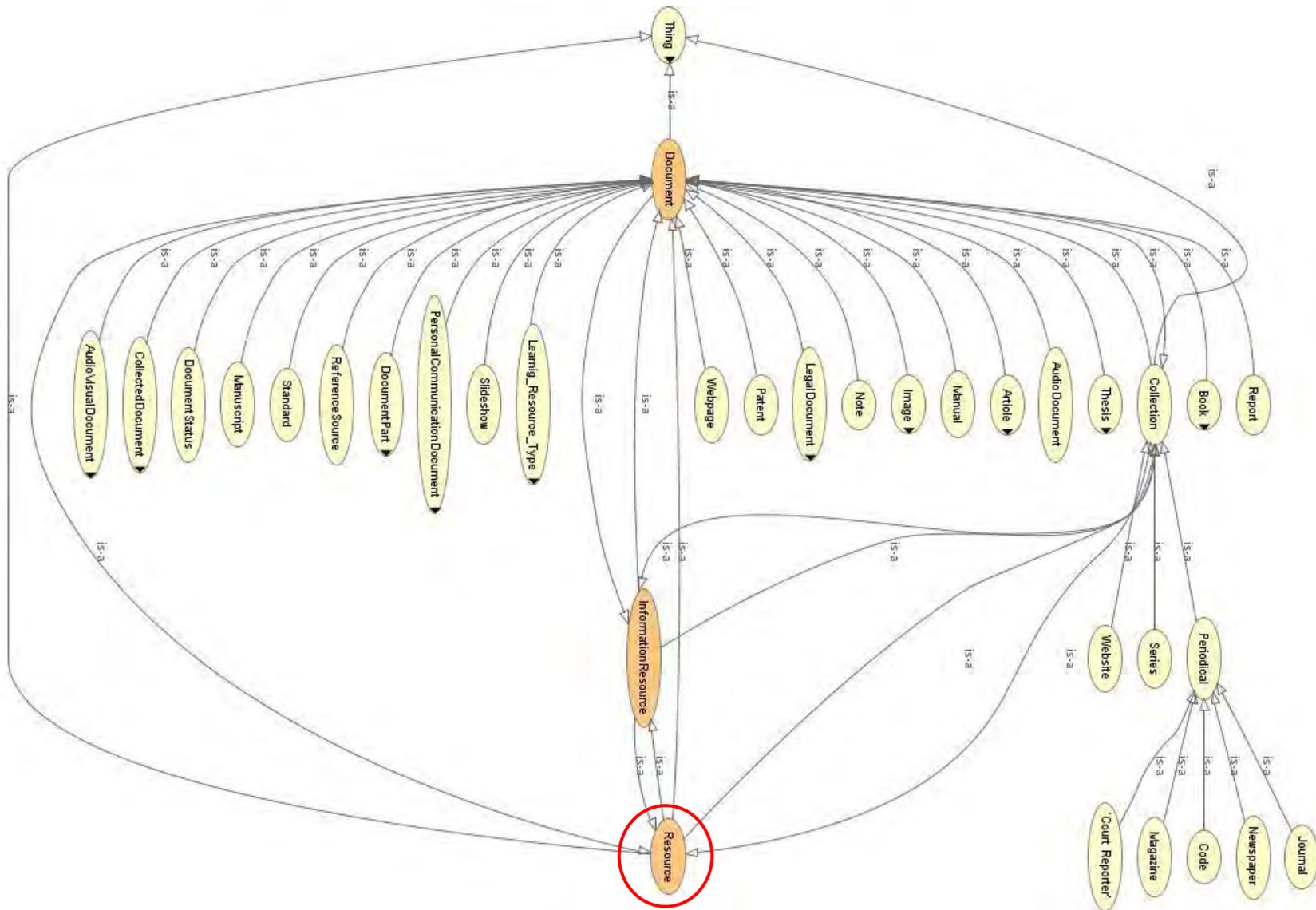


Fig. 6.26. UH ontology: Relaciones de la clase Resource

→ **Concept**

Clase asumida como unidad básica de la forma de conocimiento humana. Solo tiene relación de subordinación con Thing, al ser definida como clase principal de la ontología (ver Fig.6.27.)



Fig. 6.27. UH ontology: Relaciones de la clase Concept

→ **Credential**

Definida para el registro de certificados y números de licencia. Clase vinculada a personas, organizaciones o grupos como propietarios, responsables o beneficiarios de la misma. Contiene las subclases Certificate y License. (Ver fig.6.28)

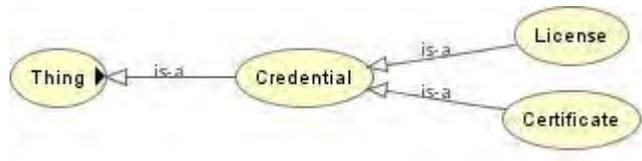


Fig. 6.28. UH ontology: Relaciones de la clase Credential

→ **EducationalTraining**

Clase referida a los estudios universitarios posteriores al título de grado, que comprende los estudios de maestría, doctorado, postdoctorado y cursos de especialización dentro de este ámbito académico. Tienen como antecedente obligatorio los estudios de pregrado, y solo se puede acceder a ellos tras la obtención del grado. (Ver Fig.6.29)



Fig. 6.29. UH ontology: Relaciones de la clase EducationalTraining

→ **Event**

Clase definida para identificar los diferentes eventos o actividades planificadas de manera periódica y que convoca a gran número de personas u organizaciones a los que pueda estar vinculada la UH como anfitrión o participante.

Event contiene varias subclases (ver en la figura 6.30 enlaces con líneas continuas). Establece relación de tipo organizer en calidad de dominio con la clase Agent y en calidad de rango con el núcleo de la ontología Thing de tipo hasProceedings. (Ver Fig.6.30).

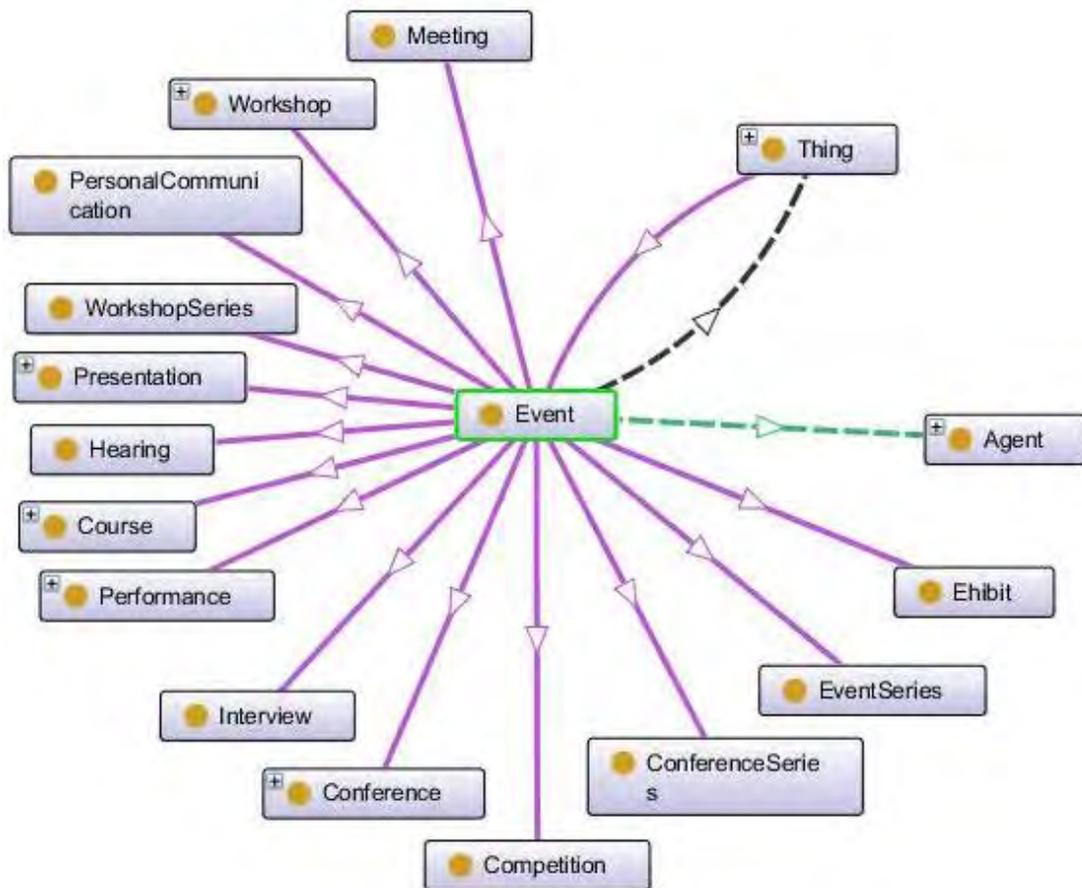


Fig. 6.30. UH ontology: Clase Event. Visualización desde Protegé

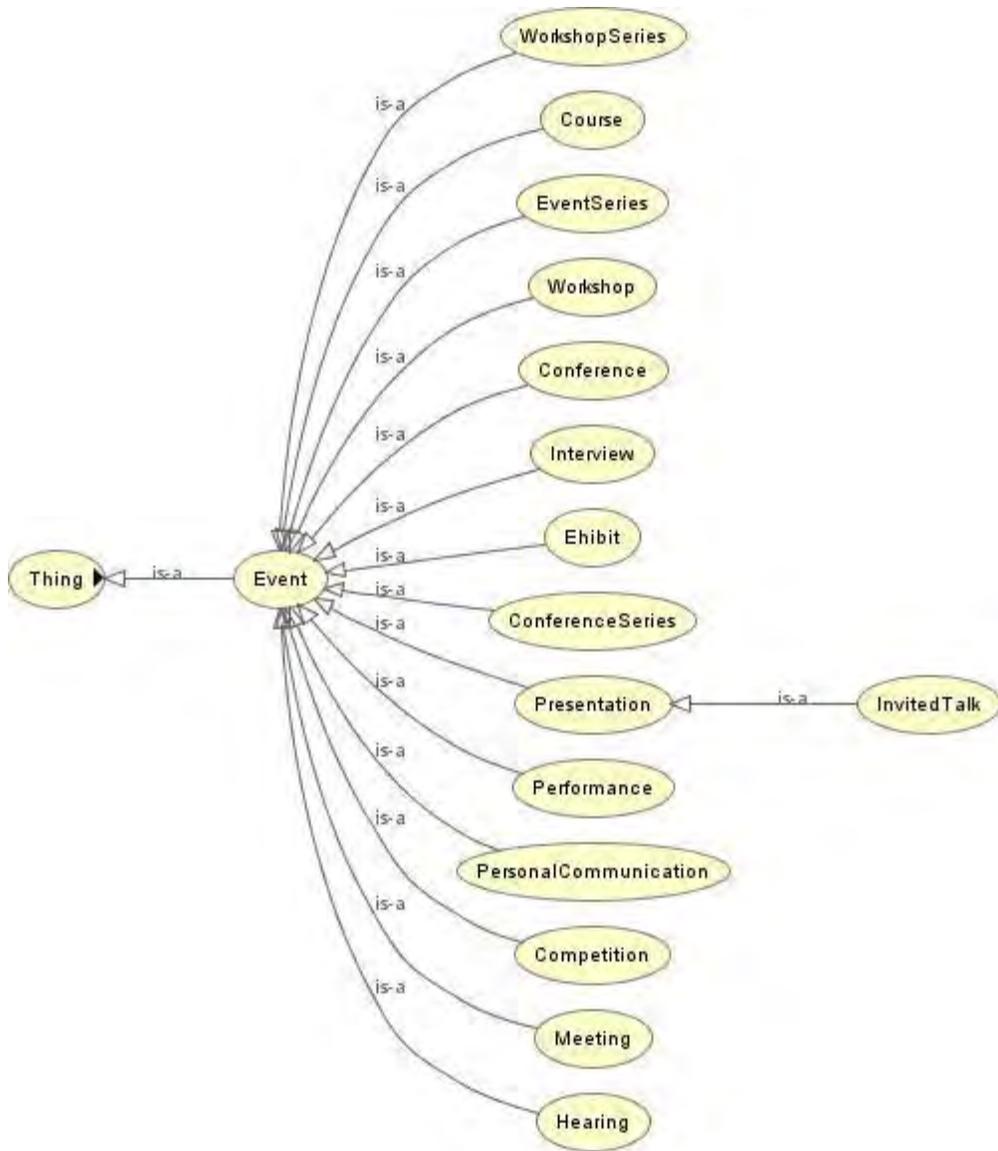


Fig. 6.31. UH ontology: Relaciones de jerarquía de la clase Event

→ **Project**

Clase definida para identificar el desarrollo de proyectos vinculados a la UH. Solo tiene relación de subordinación con Thing, al ser definida como clase principal de la ontología. (Ver fig.6.32)

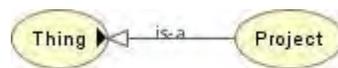


Fig. 6.32. UH ontology: Relaciones de la clase Project

→ **Software**

Clase definida para identificar el desarrollo y uso de softwares vinculados a la UH. Solo tiene relación de subordinación con Thing, al ser definida como clase principal de la ontología. (Ver Fig. 6.33)

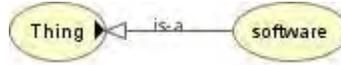


Fig. 6.33. UH ontology: Relaciones de la clase Software

→ **Area**

Clase definida para identificar áreas geográficas. Contiene a las subclases: territory y group. (Ver Fig. 6.34)

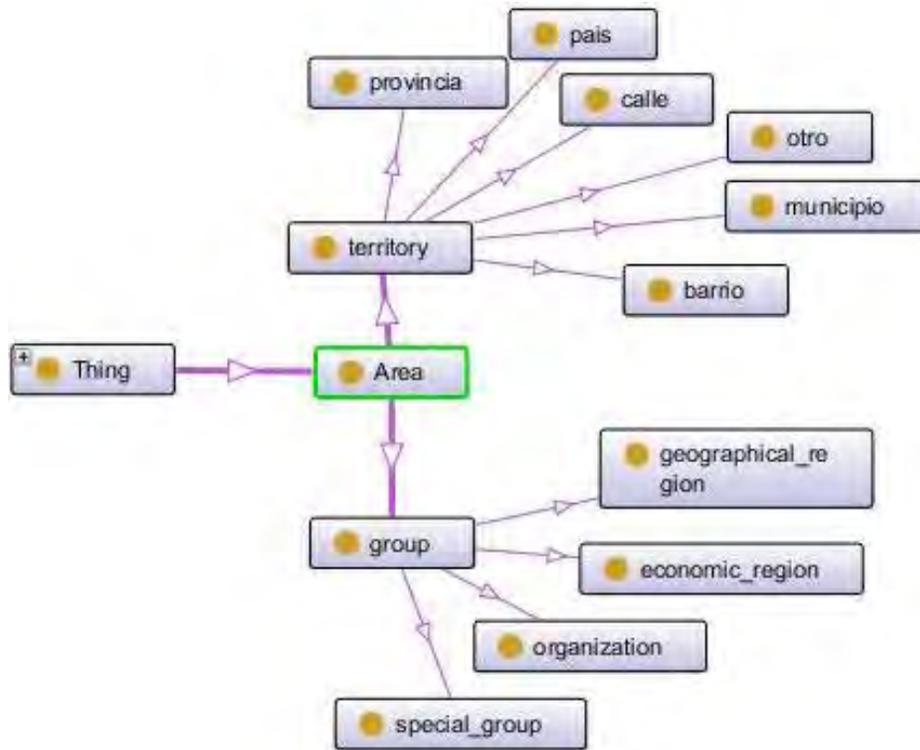


Fig. 6.34. UH ontology: Relaciones de la clase Area

6.1.5.1.3. Propiedades de los objetos

Las propiedades OWL representan relaciones entre dos objetos o individuos. Particularmente las propiedades de objetos, son establecidas desde OWL y permiten relacionar un individuo con otro (Florez Fernandez, 2013, Horridge et al., 2004).

Las propiedades de los datos definidas para UH ontology son las siguientes:

- **accessprovidedby**: “acceso ofrecido por”, establecido como propiedad para designar mención de responsabilidad sobre el acceso a la información.
- **administers**: administradores es la propiedad establecida para clarificar dicha función.
- **AdvisingRelationship**: tutor se establece como propiedad para referirse al que realiza el asesoramiento, seguimiento, guarda o custodia de personas o bienes.
- **agent**: agente se utiliza para enlazar una *persona* (un agente) a un *evento* (una conferencia, un conversatorio, etc.).
- **assignedby**: “asignado por” establecido como propiedad para designar mención de responsabilidad sobre asignación del agente.
- **assigneeforpatent**: asignado por patente.
- **assigns**: asignados.
- **awardConferred**: premio conferido.
- **base_near**: *con sede cerca* se utiliza para enlazar un agente, en relación con los recursos bibliográficos, a un lugar donde tiene su sede cerca de: puede ser una ciudad, un monumento, un edificio, etc.
- **citedBy**: *citado por* se utiliza para relacionar un documento a otro documento que cita al primer documento. Por tanto tiene definido como rango y dominio de la relación a la clase *Document*.
- **contributor**: *contribuidor* se utiliza para vincular un material bibliográfico a un colaborador: puede ser un autor, un editor, una editorial, etc. Contiene las siguientes subpropiedades:
 - **director**: referido a director de cine. Relación establecida entre la subclase *AudioVisualDocument* (de *Document*) como dominio y la clase *Agent* como rango.

- **editor:** es una persona que tiene la responsabilidad de gestión y, a veces la formulación de políticas para la parte de redacción de una editorial o de un periódico, una revista u otro tipo de publicación. Relación establecida entre la clase *Document* o *Collection* como dominio y *Agent* como rango.
 - **interviewee:** el entrevistado es un agente que es entrevistado por otro agente. Por tanto la clase *Agent* es dominio y rango a la vez.
 - **Interviewer:** un agente que realiza una entrevista a otro agente. Por tanto la clase *Agent* es dominio y rango a la vez.
 - **performer:** ejecutante es el agente que ejecuta. Por tanto se establece una relación entre la subclase *Performance* (de la clase *Event*) como dominio y la clase *Agent* como rango.
 - **translator:** el traductor es una persona que traduce el documento escrito de un idioma a otro. Relación establecida entre la clase *Document* o *Collection* como dominio y *Agent* como rango.
- **courseOfferedBy:** "curso ofrecido por" propiedad concerniente a un agente y vinculado a event.
 - **court:** una institución asociada con la expedición de un documento legal; por ejemplo, el que haya expedido una decisión. Relación establecida entre la subclase *LegalDocument* (de la clase *Document* o *Collection*) como dominio y la subclase *Organization* (de la clase *Agent*) como rango.
 - **CredentialOf:** credencial de.
 - **CurrentlyHeadOf:** "actualmente jefe de" propiedad consignada a un agente (person) en relación con otro agente (person, organization o group).
 - **degree:** no se define mediante una enumeración, sino como se define el rango de bibo: degree en la lista definida de Bibo: ThesisDegree. Se espera que el personal que interactúa con los datos de la ontología sea capaz de definir nuevos Degree si es necesario por algunos casos de uso especiales. La creación de dicha enumeración restringiría que esto suceda. Es la relación establecida entre

la subclase *Thesis* (de la clase *Document* o *Collection*) como dominio y *Thesis degree* como rango.

- **degreeEarned**: título obtenido.
- **distributor**: distribuidor de un documento o una colección de documentos. Relación establecida entre la clase *Document* o *Collection* como dominio y *Agent* como rango.
- **DomesticGeographyFocus**: enfoque geográfico nacional.
- **editor**: la propiedad "editor" referida a la clase *Agent* se utiliza para identificar a la persona que tiene por oficio editar libros, periódicos, películas, discos, entre otros. Se utiliza para relacionar a *Agent* con *Document*.
- **EducationalTrainingOf**: referido a las acciones de superación postgraduadas.
- **EventWithin**: definida para establecer actividades y acciones realizadas en el marco de un evento.
- **Format**: *formato* se utiliza para describir el formato de un recurso bibliográfico.
- **grantAwardedBy**: subvención concedida por.
- **hasAcademicDegree**: para designar el grado académico de una persona.
- **hasCredential**: tiene credencial o autorización para. Es una propiedad consignada para la clase *Agent*.
- **hasCurrentMember**: para consignar los miembros de la subclase *Organización* (grupo, organización o institución) o la clase *Event*.
- **hasEditorRole**: tiene role de editor.
- **hasInvestigatorRole**: tiene rol de investigador.
- **haspart**: es parte de.
- **hasPrerequisite**: tiene prerequisite.

- **hasPrincipalInvestigatorRole**: es primer investigador (para el dominio de la clase *Person*).
- **hasProceedings**: tiene procedimientos (para el dominio de la clase *Event*).
- **hasResearchArea**: tiene área de investigación.
- **hasReviewerRole**: tiene rol de revisión.
- **hasServiceProviderRole**: tiene rol de proveedor de servicios.
- **hasTeacherRole**: tiene rol docente.
- **hisTopicIS**: su tema es.
- **homepage**: se utiliza para unir un agente a su página de inicio (que es una página web accesible a través de una URL).
- **IncludesEvent**: incluye eventos.
- **inEventSeries**: en serie de eventos.
- **isPartOf**: *es parte de* , contiene a la subpropiedad *reproducedIn*.

reproducedIn se utiliza para representar la relación del recurso en el que se produce otro recurso. Tiene dominio y rango en la clase *Document*.

- **isReferencedBy**: se utiliza para relacionar una cita de referencia a un recurso bibliográfico (para el dominio de la clase *Person*). Contiene la subpropiedad *subsequentLegalDecision* (para registrar la decisión legal en la apelación que actúa en un caso (afirmarlo, darle la vuelta, etc.)). Con dominio y rango de la subclase *Decision* de la clase *LegalDocument*.
- **isVersionOf**: *es versión de* (para el dominio de la clase *Document*), contiene la subpropiedad *translation of* (referido a un documento traducido del documento original. Rango y dominio de la clase *Document*).
- **lenguaje**: se utiliza para vincular un recurso bibliográfico a la lengua utilizada para expresarlo.
- **list of contributors**: es una lista ordenada de los contribuyentes, generalmente tienen en cuenta el orden de importancia. Definido para la clase *Document*. Contiene las siguientes subpropiedades:

- *list of authors*: lista ordenada de los autores.
- *list of editors*: Es una lista ordenada de los editores
- **offersDegree**: ofertas de Grado.
- **organizer**: el organizador de un evento; incluye organizadores de la conferencia, sino también a las agencias gubernamentales u otros organismos que son responsables de la realización de las audiencias. Tiene por dominio a la clase *Event* y por rango a la clase *Agent*.
- **owner**: propietario de un documento o una colección de documentos. Tiene por dominio a la clase *Document* o la clase *Collection*.
- **PersonInPosition**: situación de la persona.
- **place**: se utiliza para relacionar un evento como una conferencia con el lugar geográfico donde ocurre, por ejemplo, de París.
- **prerequisiteFor**: pre requisito para.
- **produced_in**: *producido en* contiene la subpropiedad *presented at*. Dicha subpropiedad se refiere al documento de un evento; por ejemplo, la ponencia para una conferencia.
- **producer**: productor de un documento o una colección de documentos. Tipo de relación establecida entre la clase *Document* o *Collection* como dominio y *Agent* como rango.
- **product**: se utiliza para enlazar un evento (como una conferencia) para un resultado (un producto) de ese evento, por ejemplo, un artículo, un procedimiento, etc. Contiene la subpropiedad *presented at* que relaciona un evento a los documentos asociados.
- **publisher**: se utiliza para vincular un material bibliográfico a su editor. Contiene la subpropiedad *issuer* como encabezada de emitir los documentos publicados, a menudo informalmente, como comunicados de prensa, informes, etc.

La subpropiedad *issuer* (o emisor) establece una relación entre la clase *Document* o *Collection* como dominio y *Agent* como rango.

- **receipt:** recibido.
- **recipient:** referido a un agente que recibe un documento de comunicación. Por tanto establece una relación entre la subclase *PersonalCommunicationDocument* (perteneciente a la clase *Document*) como dominio y la clase *Agent* como rango.
- **references:** referencia contiene la subpropiedad *cites*, que relaciona un documento a otro que se cita por el primero como referencia, comentario, opinión, cita o para otro propósito. Por tanto establece una relación como dominio y rango con la clase *Document*.
- **relation:** La propiedad *relation* (relación) contiene la siguiente subpropiedad:
 - **annotates:** es la subpropiedad para establecer una nota crítica o explicativa para un documento. Establece una relación entre la subclase *Note* (perteneciente a la clase *Document*) como dominio y la clase *Resource* como rango
 - **review of:** se relaciona con un documento de opinión sobre una cosa (de recursos, artículo, etc.). Establece una relación entre la clase *Document* como dominio y la clase *Resource* como rango.
 - **transcript of:** se refiere a un documento de un original transcripto. Establece una relación entre la clase *Document* como dominio y la clase *Resource* como rango.
- **rights:** se utiliza para describir los derechos relacionados con un recurso bibliográfico.
- **sponsoredBy:** patrocinado por es una propiedad consignada para la clase *Agent*
- **status:** se emplea igual que el rango de *bibo:status*, definido en la lista de *bibo:DocumentStatus*. Se espera que el personal que interactúa con los datos de la ontología sea capaz de definir nueva estatus si es necesario por algunos casos de uso especiales como, por ejemplo, el estado de la publicación de contenidos (normalmente académica).

- **sub_event**: se utiliza para vincular los grandes eventos con los eventos más pequeños, como talleres que se producen en el contexto de una conferencia.
- **SubGrantOf**: subtítulo de un Document
- **Subject**: se utiliza para describir la temática de un recurso bibliográfico.
- **time**: se utiliza para describir el tiempo de duración un evento o acción.
- **title**: se utiliza para describir el título de un recurso bibliográfico
- **value**: se utiliza para describir el contenido de un "Bibo:Document" y de otros recursos bibliográficos. Se aconseja utilizar esta propiedad en lugar de la "Bibo:content".

6.1.5.1.4. Propiedades de los datos

Las propiedades OWL representan relaciones entre dos objetos o individuos. Particularmente las propiedades de datos, son establecidas desde OWL y permiten relacionar un individuo con un XML Schema Data-type value o un literal RDF (Florez Fernandez, 2013, Horridge et al., 2004).

Las propiedades de los datos definidas para UH ontology son las siguientes:

- **abstract**: con el propósito de acceder al resumen de los recursos.
- **changeNote**: para consignar cambio en una nota.
- **countryArea**: área o país.
- **date**: se utiliza para vincular un material bibliográfico a la fecha de un evento. Compruebe dcterms:created y other para especializaciones, adecuados para esta propiedad. Esta propiedad contiene las subpropiedades: *created* (representa la creación de un item bibliográfico) y *issued* (representa el dato numérico en un recurso bibliográfico).
- **date argued**: la fecha en la cual un caso legal se alega antes de un tribunal. El formato de la fecha es xsd: date.
- **description**: usado para describir un recurso bibliográfico.
- **doi**: representa la ubicación física de un documento antes de ser publicado.

- **edition**: edición.
- **editorialNote**: nota editorial.
- **elssn**: es equivalente a *eissn*.
- **endingPage**: es equivalente a *'page end'*.
- **family_name**: equivalente al apellido de familia de una persona cuando guarda relación con un recurso bibliográfico.
- **givenname**: equivalente al nombre de pila de una persona relacionada con un recurso bibliográfico.
- **identifier**: es un identificador. Esta propiedad contiene a las subpropiedades: *asin, coden, doi, eanucc13, eissn, isbn, issn, lccn, oclnum, pmid, sici, upc* y *uri*.
- **isbn**: para el número de identificación internacional asignado a los libros.
- **issn**: para el número de identificación internacional asignado a las publicaciones periódicas.
- **issue**: el término issue se atribuye a la unidad de trabajo para realizar una mejora en un Sistema informático. Un issue puede ser el arreglo de un fallo, una característica pedida, una tarea, un pedido de Documentación específico y todo tipo de solicitud al equipo de desarrollo.
- **localityName**: se utiliza como número local para nombrar a la localidad de un editor, y autor, etc.
- **locator**: una descripción (para este caso numérica) que localiza un elemento dentro de un documento o colección que contiene. Es equivalente a la propiedad *number*. Incluye como subpropiedades *chapter, issue, page end, page start, pages, section, volume*.
- **number**: equivalente a *locator*.
- **numberID**: se utiliza para registrar número de carnet de identidad
- **number of pages**: se utiliza para representar el número de páginas contenidas en un documento.

- **number of volumes:** representa el número de volúmenes contenidos en una colección de documentos (por lo general una serie, publicación periódica, etc.).
- **prefix name:** referido al prefijo de un nombre.
- **PublishingGroup:** para registrar el grupo de publicación
- **short title:** título abreviado de un documento.
- **short description:** para registrar el resumen (especialmente el indicativo sobre el contenido del documento).
- **startingPage:** se utiliza para puntualizar la página de inicio.
- **suffix name:** sufijo del nombre.
- **volume:** volume.

6.1.5.1.5. Anotaciones

Las anotaciones constituyen otro tipo de propiedad. Se definen y utilizan para agregar comentarios y descripciones en lenguaje natural.

Las anotaciones previstas desde UH ontology permitirán realizar anotaciones a las clases, propiedades de objetos y propiedades de datos sobre los siguientes aspectos: changeNote, comment, creator, deprecated, description, editorialNote, example, historyNote, identifier, incompatibleWith, isDefinedBy, isPartOf, label, note, priorVersion, scopeNote, seeAlso, term_status, title y versionInfo. Además permite anotarlos identificando los siguientes idiomas: inglés (en), español (es) y francés (fr). (Ver fig.6.35)

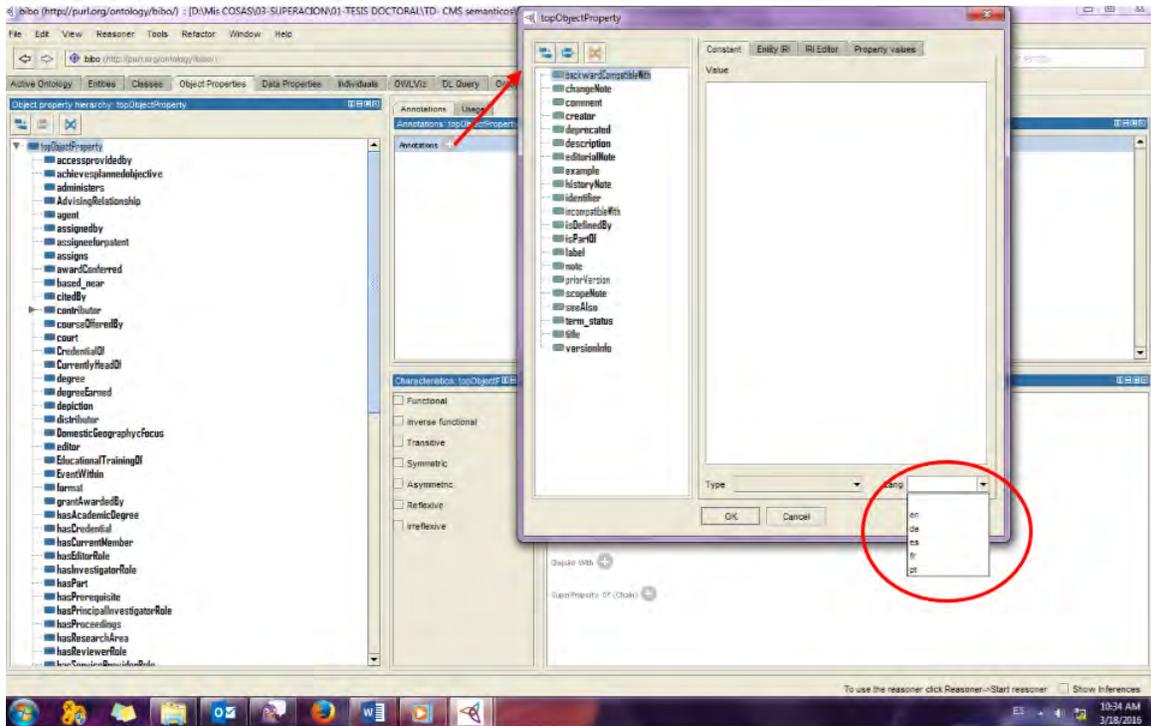
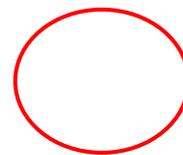


Fig.6.35. UH ontology: Sistema de anotación. Visualización desde Protege con adaptación propia

De esta manera desde la base de la ontología comienzan a agregarse anotaciones. Un ejemplo de ello lo constituyen las anotaciones incorporadas a UH ontology. (Ver Fig.6.36)



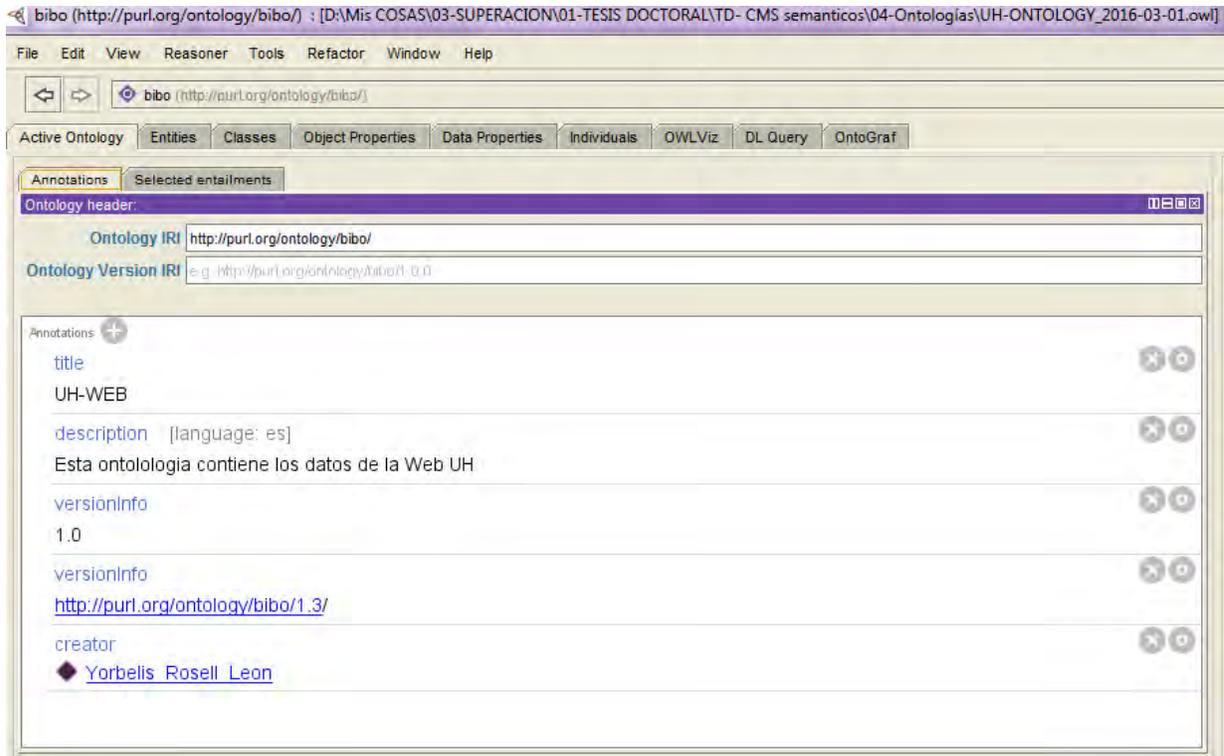


Fig. 6.36. UH ontology: Anotaciones de la ontología. Visualización desde Protegé

6.1.5.1.6. Relaciones de cardinalidad y ordenabilidad

Las restricciones de cardinalidad y ordinalidad establecidas son de tipo: (ver ejemplo en la Fig. 6.37)

- Existencial (Some)
- Universal (Only)
- Cardinalidad mínima (Min)
- Cardinalidad exacta (Exactly)
- Cardinalidad máxima (Max)

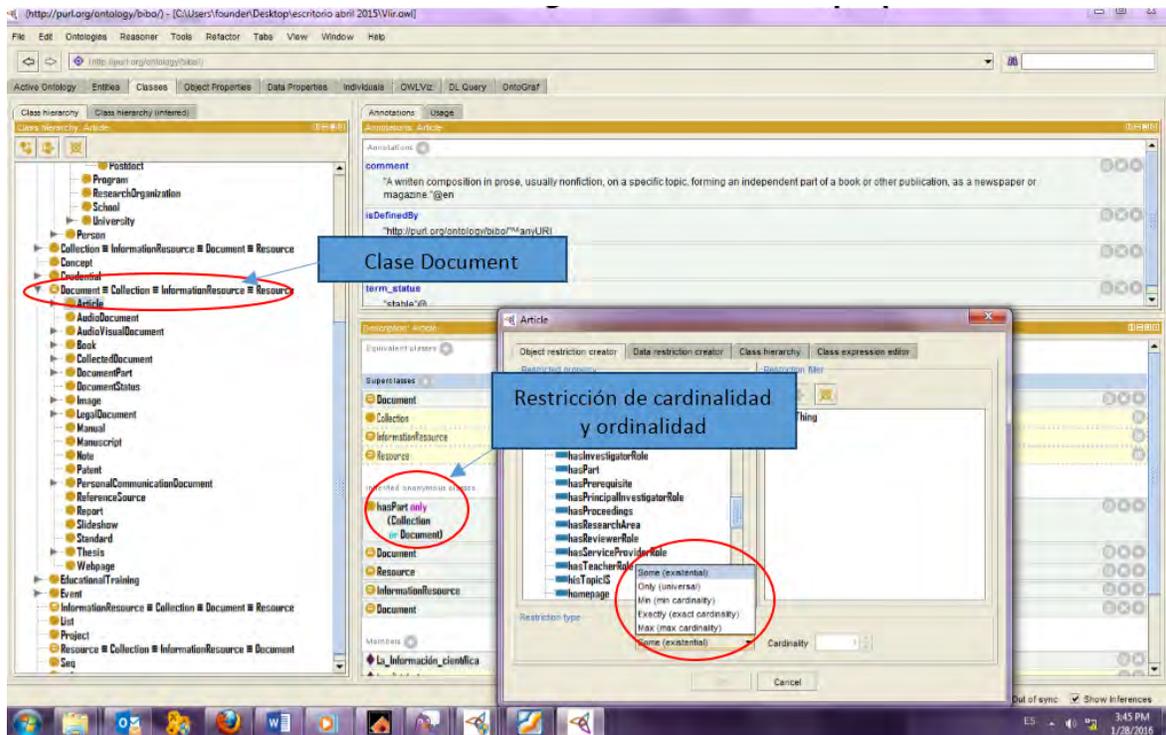


Fig.6.37. UH ontology: Ejemplo de relaciones de cardinalidad desde la Clase Document. Visualización desde Protegè con anotaciones propias

6.1.5.1.7. Gestión del conocimiento

Bajo estas características la ontología permite que, desde el CMS y los diferentes productos que se construyen con él, se desarrolle de manera coherente la gestión del conocimiento generado en la UH. (Véase la fig.6.38, que se muestra a continuación)

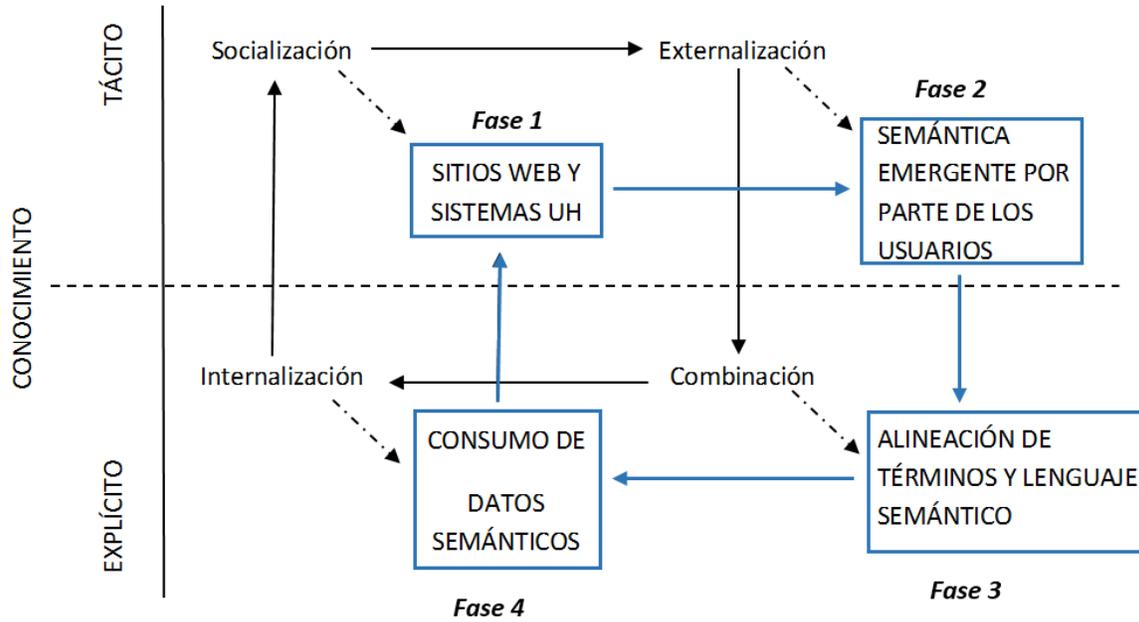


Fig. 6.38. Gestión del conocimiento desde UH ontology (elaboración propia)

Tomando como base el esquema propuesto por Nonaka y Takeuchi (1999) del ciclo de conocimiento, se modela el proceso de gestión del conocimiento (Fig.6.38) desde el manejo de los datos en la ontología propuesta.

En UH ontology se desarrollan las siguientes etapas para la gestión del conocimiento:

→ **Fase 1**: Parte la etapa de Socialización (Nonaka et al., 1999) donde se da la diseminación del conocimiento a través de los sitios web y otros sistemas de la UH, implementados desde el CMS UH-WEB. Se asume desde la socialización del conocimiento tácito desde la perspectiva del software (no se encuentra semánticamente reconocido) intentando hacer una analogía con lo que para los seres humanos sería el compartir ideas y experiencias no documentadas, pero comunicadas.

→ **Fase 2**: Parte de la etapa de Externalización (Nonaka et al., 1999), donde se transforma el conocimiento en conceptos entendibles por el software. Como herramienta la ontología permite semantizar lo que los usuarios aporten. Predominan técnicas deductivas e inductivas.

→ **Fase 3**: Parte de la etapa de Combinación (Nonaka et al., 1999). Se trabaja sobre la alineación y similitud semántica de los términos en la ontología. Se

realiza la identificación, captura e integración de nuevo conocimiento explícito esencial mediante la asignación de metadatos semánticos en el procesado de documentos, planos, comentarios, etc.

→ **Fase 4:** Parte de la etapa de Internalización (Nonaka et al., 1999), referida al enriquecimiento del conocimiento tácito del usuario (semánticamente identificado y/o representado) a partir de la consulta y asimilación del conocimiento explícito (ontología).

Es un proceso cíclico que permitirá enriquecimiento y gestión del conocimiento generado desde la UH.

6.1.5.2. Lenguajes Ontológicos utilizados

Los lenguajes ontológicos que se utilizan en el sistema son diversos y están en función de lograr la calidad de las descripciones y anotaciones semánticas. A continuación se muestran los lenguajes ontológicos empleados en el sistema y las ontologías reutilizadas para hacer más rápido el proceso de diseño:

✓ **SKOS** (W3C, 2012b) : Organiza las materias a partir de la clase SKOS:Concept e incluye las propiedades *skos:prefLabel* y *Skos:altLabel*, con las que se logran las relaciones sindéticas entre términos, en el diseño se suprimieron las relaciones de mapeo ya que se desarrollaron otras propiedades particulares dentro del CMS para mapear los léxicos. Se ha decidido usar el esquema del tesauruso de la Unesco para normalizar la actividad de la universidad. Las relaciones definidas a nivel ontológico pueden verse en la figura siguiente (ver figura 6.39)

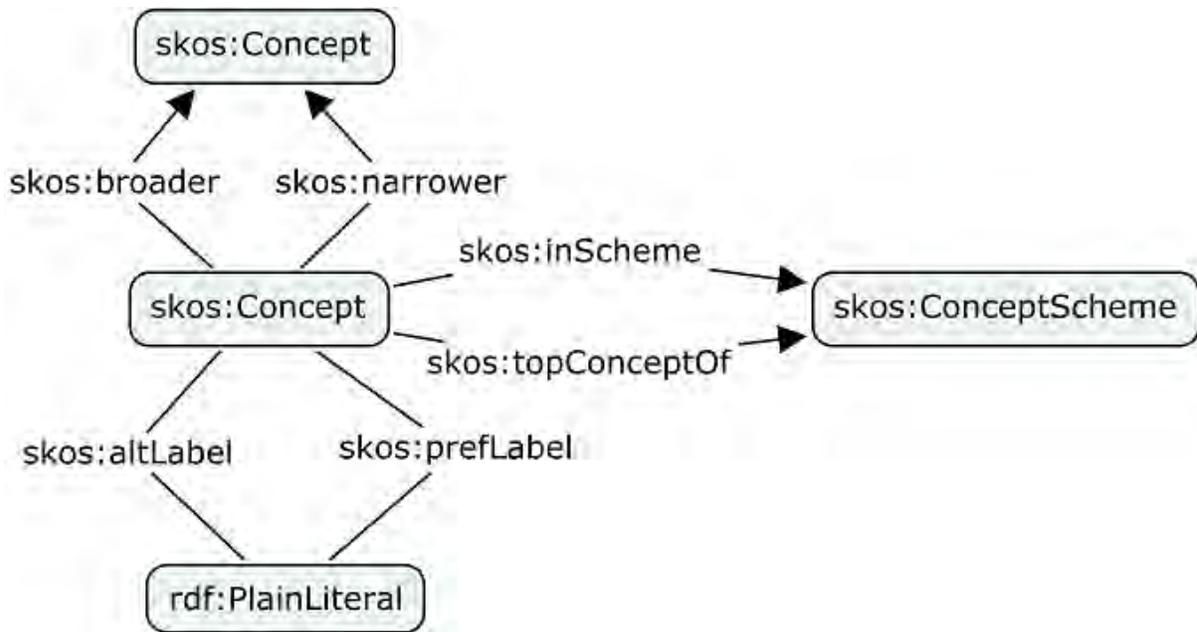


Fig. 6.39. Clases fundamentales de SKOS (Rensselaer-Polytechnic-Institute and Tetherless-World-Constellation, 2014)

✓ **Bibo** (D'Arcus and Giasson, 2009): Es una ontología que se ajusta a la descripción que se precisaría para los documentos que existen en la Universidad de La Habana y se conecta con las clases de **FOAF** a través de otras clases entre las que se encuentran: **Colecciones, Recursos de Información, Documentos**. Las mismas han sido unidas en el modelo ontológico formando una restricción que las conecta como clases equivalentes que describen: artículos, documentos audiovisuales, libros, proceeding, parte de documentos, imágenes, documentos legales, manuales, manuscritos, notas, patentes, comunicaciones personales y sus referencias. (Ver figura 6.40 que se muestra a continuación). Las clases de **bibo** al estar asociadas a las de **foaf** y permiten la realización de anotaciones en formato **Dublin Core**.

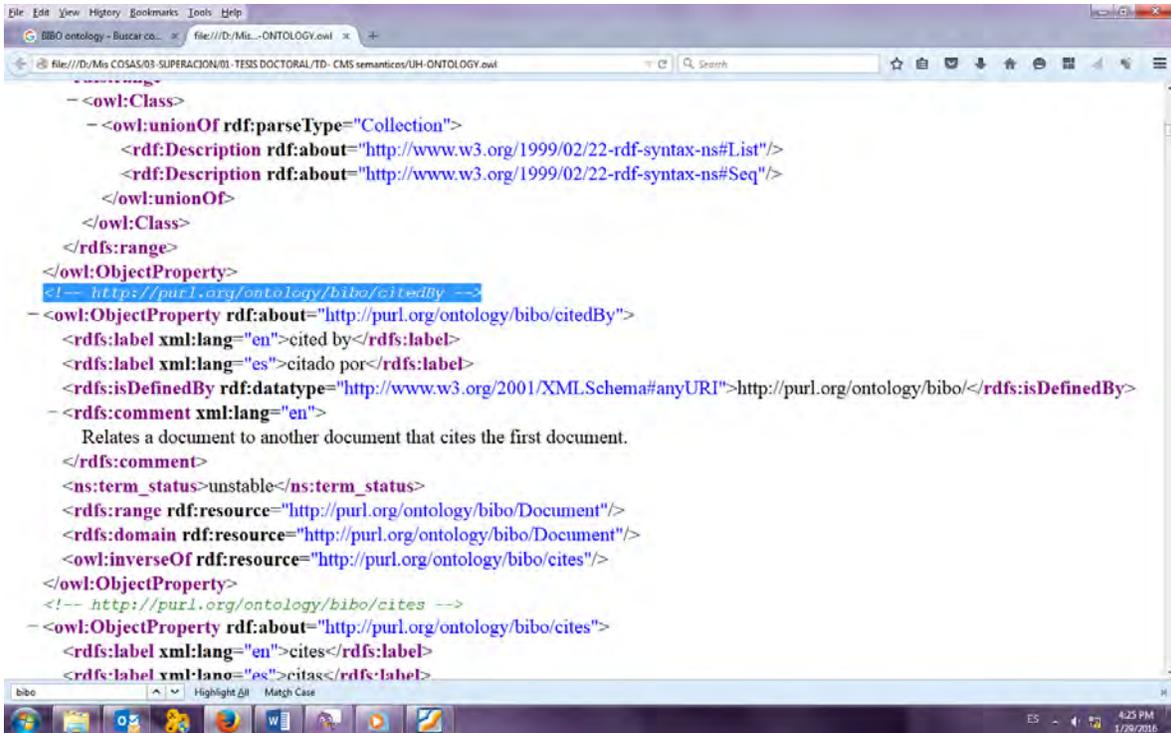


Fig. 6.40. Vista HTML del uso de BIBO en UH ontology

✓ **VIVO** (Cornell, 2009): Es una ontología que se nutre de relaciones con otros vocabularios ontológicos. Su objetivo fundamental está en crear un CRIS con los datos de investigación de los docentes y estudiantes de la Universidad de La Habana. (Ver fig. 6.41)

```

- <!--
  http://vivoweb.org/ontology/core#AdvisingRelationship
-->
- <owl:ObjectProperty rdf:about="http://vivoweb.org/ontology/core#AdvisingRelationship">
  <rdf:label xml:lang="es">Tutor</rdf:label>
</owl:ObjectProperty>
<!-- http://vivoweb.org/ontology/core#CredentialOf -->
- <owl:ObjectProperty rdf:about="http://vivoweb.org/ontology/core#CredentialOf">
  <rdf:label xml:lang="es">Credencial de</rdf:label>
</owl:ObjectProperty>
<!-- http://vivoweb.org/ontology/core#CurrentlyHeadOf -->
- <owl:ObjectProperty rdf:about="http://vivoweb.org/ontology/core#CurrentlyHeadOf">
  <rdf:label xml:lang="es">actualmente jefe de</rdf:label>
</owl:ObjectProperty>

```

Fig.6.40. Vista HTML de UH Ontology. Ejemplificación de VIVO

Las clases en VIVO se relacionan con diversas instancias identificadas con una URI base única. Las instancias de la clase son individuales y se identifican de forma única por cada grado persona o actividad. Para lograr relaciones eficientes en vivo se utiliza la clase ISF, la que tiene la intención de relacionar mediante restricciones aspectos de la vida académica de un docente ejemplo: relaciona un grado académico único con la organización que lo emite, la persona que lo recibe, los valores de tiempo, ubicación, etc. (que es una representación de una relación n-aria). Además las instancias de esta clase también pueden especificar el tipo de título recibido. La opción desarrollada por la autora ha sido crear clases adicionales en su modelo ontológico para separar y definir de la forma más coherente los premios, distinciones y reconocimientos de la Universidad de la Habana incluyendo clases propias. En la figura 6.41 y/o 6.42, se puede observar su interacción con las clases de otros vocabularios ontológicos como: bibo, geontology, foaf.

```

<rdfs:subPropertyOf rdf:resource="http://purl.org/dc/terms/contributor"/>
<rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
- <rdfs:domain>
  - <owl:Class>
    - <owl:unionOf rdf:parseType="Collection">
      <rdf:Description rdf:about="http://purl.org/ontology/bibo/Collection"/>
      <rdf:Description rdf:about="http://purl.org/ontology/bibo/Document"/>
    </owl:unionOf>
  </owl:Class>
</rdfs:domain>
</owl:ObjectProperty>

```

Fig. 6.42. Vista HTML de UH Ontology. Ejemplificación de VIVO y su interacción con las clases de otros vocabularios ontológicos

En el diseño de la ontología del CMS se incluyó una clase **Event** que viene del vocabulario para agrupar los eventos descritos por *vivo-ontology*. (Ver ejemplo en la fig.6.43). La misma sirve de clase de agrupamiento a las subclases: Exhibición, Performance, Eventos seriados, Conferencias, Reuniones, Cursos, Presentaciones, Comunicaciones Personales, Talleres, Competencias, Entrevistas. (Ver figura. 6.43).



Fig. 6.43. Clase Event y manejo de sus propiedades (elaboración propia desde la UH ontology visualizada en PROTEGE)



Fig. 6.44. Visualización de la clase Event desde la ontología. Vista HTML

Las propiedades y sentencias describiendo conceptos jerárquicos de la clase eventde la ontología de eventos es la siguiente: factor, agent, time, factor, sub-event (ver fig. 6.45)

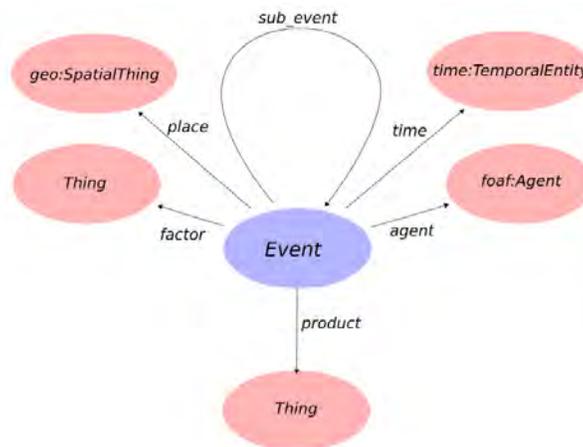


Fig.6.45.Modelo de la clase Event (Raimond and Abdallah, 2007)

✓ **Goentología:** La geo-ontología tiene como base a la clase área de VIVO (ver fig. 6.46) y se usa en el sistema se desarrollado a partir de una metodología

para la representación de datos Ontología de Representación de Datos (ORD) propuesta por Larin-Fonseca y Garea Llano (Larin-Fonseca and Garea-Llano, 2011) en la cual la integración de datos semánticos es la clave (ver fig. 6.47).

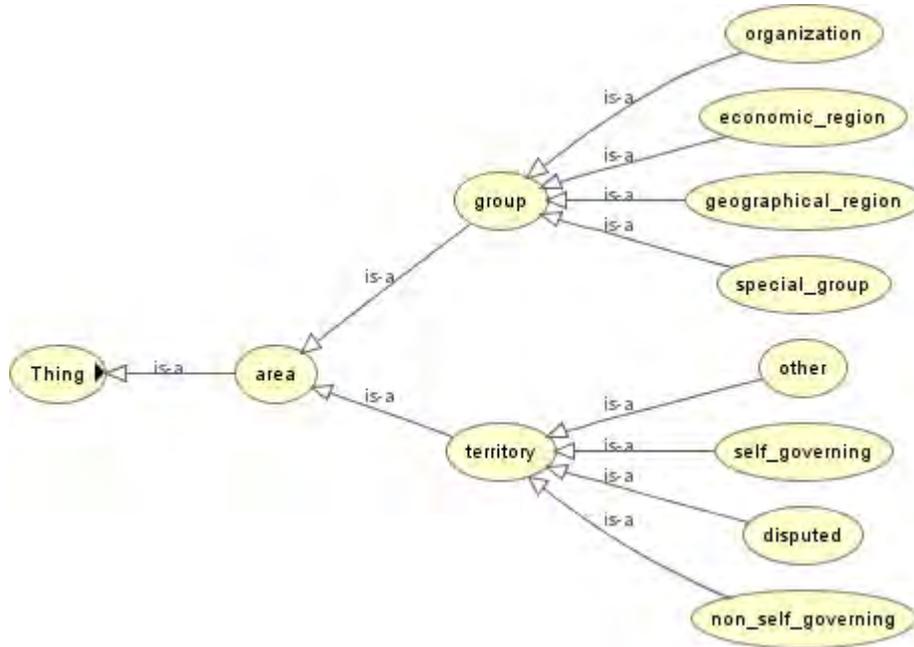


Fig. 6.46. Clase Area y manejo de sus propiedades (elaboración propia desde la UH ontology visualizada en PROTEGE)

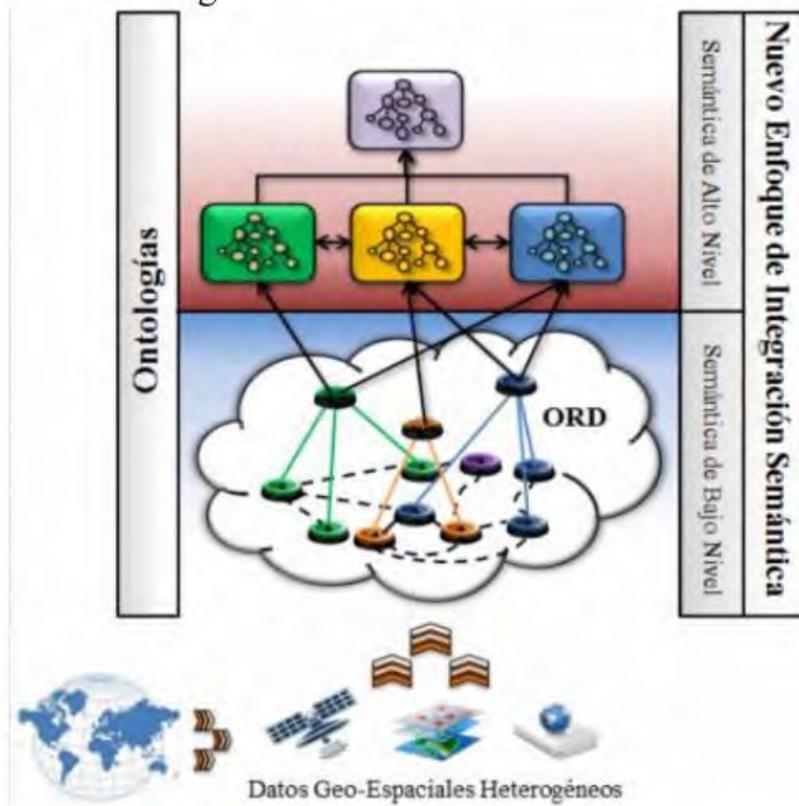


Fig. 6.47. Enfoque de integración semántica de datos geoespaciales (Larin-Fonseca and Garea-Llano, 2011)

Este tipo de ontología provee un mayor enriquecimiento semántico en el proceso de integración ya que permite representación profunda de los contenidos. El enfoque que se han seguido aquí es el de ontología única que usa información de fuentes heterogéneas.

La estructura flexible y dinámica de la ORD permite capturar las características de los objetos geoespaciales descritas en los datos y las relaciones que se generan entre ellos. Esta propuesta es realizada sobre la premisa de que la integración de los datos geo-espaciales ha de realizarse tomando en cuenta la naturaleza semántica que subyace en este tipo de datos en combinación con las Ontologías de Niveles Superiores (ONS).

La fuente de datos utilizada es el tesoro de la FAO (AGROVOC) que posee especificaciones para datos geográficos explícitos, lo que no entra en contradicción con el modelo de datos que se propone.

Para lograr este objetivo se propone el uso de la ORD para la representación explícita de la naturaleza semántica de los geo-datos y las relaciones entre ellos y posteriormente su vinculación con las diferentes ONS en dependencia del enfoque de integración semántica a alto nivel utilizado. Entre las principales ventajas del uso de esta nueva extensión podemos mencionar que:

- a) Se habilita la integración tomando en cuenta la semántica de bajo nivel.
- b) La incorporación del conocimiento subyacente en los geo-datos y las relaciones entre ellos en el proceso de integración.
- c) Permite el uso de diferentes enfoques de integración semántica a alto nivel incluso simultáneamente.
- d) Habilita la integración basada en representaciones semánticas multidimensionales de los objetos.
- e) Habilita el trabajo con fuentes de datos dinámicas y heterogéneas.
- f) Habilita la posible generación de nuevos niveles de abstracción.

✓ **DublinCore:** Este es la ontología que hemos usado para la anotación manual y automática del contenido, se basa en RDF y facilita la interoperabilidad entre aplicaciones. Dublin Core es un sistema de metadatos que centra su extensión en el manejo de 15 definiciones semánticas descriptivas que pretenden transmitir un significado semántico y físico de los documentos. Su extensión y aplicabilidad se basa en la capacidad de proporcionar la información descriptiva básica sobre cualquier recurso, sin que importe el formato de origen, el área de especialización o el origen cultural. Sus elementos se clasifican en tres grupos:

- Elementos relacionados principalmente con el contenido del recurso.
- Elementos relacionados principalmente con el recurso cuando es visto como una propiedad intelectual.
- Elementos relacionados principalmente con la instanciación del recurso

En el sistema de anotación se han usado los siguientes elementos:

1. **Título:** el nombre dado a un recurso, habitualmente por el autor (Etiqueta: DC.Title)

- **Claves:** los temas del recurso. Típicamente, Subject expresará las claves o frases que describen el título o el contenido del recurso. Se fomentará el uso de vocabularios controlados y de sistemas de clasificación formales. Etiqueta: DC.Subject

- **Descripción:** una descripción textual del recurso. Puede ser un resumen en el caso de un documento o una descripción del contenido en el caso de un documento visual. Etiqueta: DC.Description

- **Fuente:** secuencia de caracteres usados para identificar unívocamente un trabajo a partir del cual proviene el recurso actual. Etiqueta: DC.Source

- **Relación:** es un identificador de un segundo recurso y su relación con el recurso actual. Este elemento permite enlazar los recursos relacionados y las descripciones de los recursos. Etiqueta: DC.Relation

- **Autor o Creador:** la persona u organización responsable de la creación del contenido intelectual del recurso. Por ejemplo, los autores en el caso de documentos escritos; artistas, fotógrafos e ilustradores en el caso de recursos visuales. Etiqueta: DC.Creator

- **Editor:** la entidad responsable de hacer que el recurso se encuentre disponible en la red en su formato actual. Etiqueta: DC.Publisher

- **Otros Colaboradores:** una persona u organización que haya tenido una contribución intelectual significativa, pero que esta sea secundaria en comparación con las de las personas u organizaciones especificadas en el elemento Creator. (por ejemplo: editor, ilustrador y traductor). Etiqueta: DC.Contributor

- **Derechos:** son una referencia (por ejemplo, una URL) para una nota sobre derechos de autor, para un servicio de gestión de derechos o para un servicio que dará información sobre términos y condiciones de acceso a un recurso. Etiqueta: DC.Rights

- **Fecha:** una fecha en la cual el recurso se puso a disposición del usuario en su forma actual. Esta fecha no se tiene que confundir con la que pertenece al elemento Coverage, que estaría asociada con el recurso en la medida que el contenido intelectual está de alguna manera relacionado con aquella fecha. Etiqueta: DC.Dat

- **Formato:** es el formato de datos de un recurso, usado para identificar el software y, posiblemente, el hardware que se necesitaría para mostrar el recurso. Etiqueta: DC.Format

- **Identificador del Recurso:** secuencia de caracteres utilizados para identificar unívocamente un recurso. Ejemplos para recursos en línea pueden ser URLs y URNs. Para otros recursos pueden ser usados otros formatos de identificadores, como por ejemplo ISBN ("International Standard Book Number"). Etiqueta: DC.Identifier.

- ✓ **FOAF:** FOAF es el acrónimo de una ontología *Friend of a Friend* que posee diversas clases entre ellas: Agent, utilizada para la definición de 3 entidades básicas; Group, Organization y Person (Ver fig. 6.48), las mismas facilitan la organización de todos los miembros de la comunidad universitaria, las organizaciones que contiene la organización que contiene y con las que se relaciona, además de los grupos de trabajo. Las propiedades de los que facilitan la concepción lógica de la propuesta son: foaf:mbox, foaf:homepage, foaf:surname, foaf:givenname, foaf:birthaday , foaf:knows, foaf:title

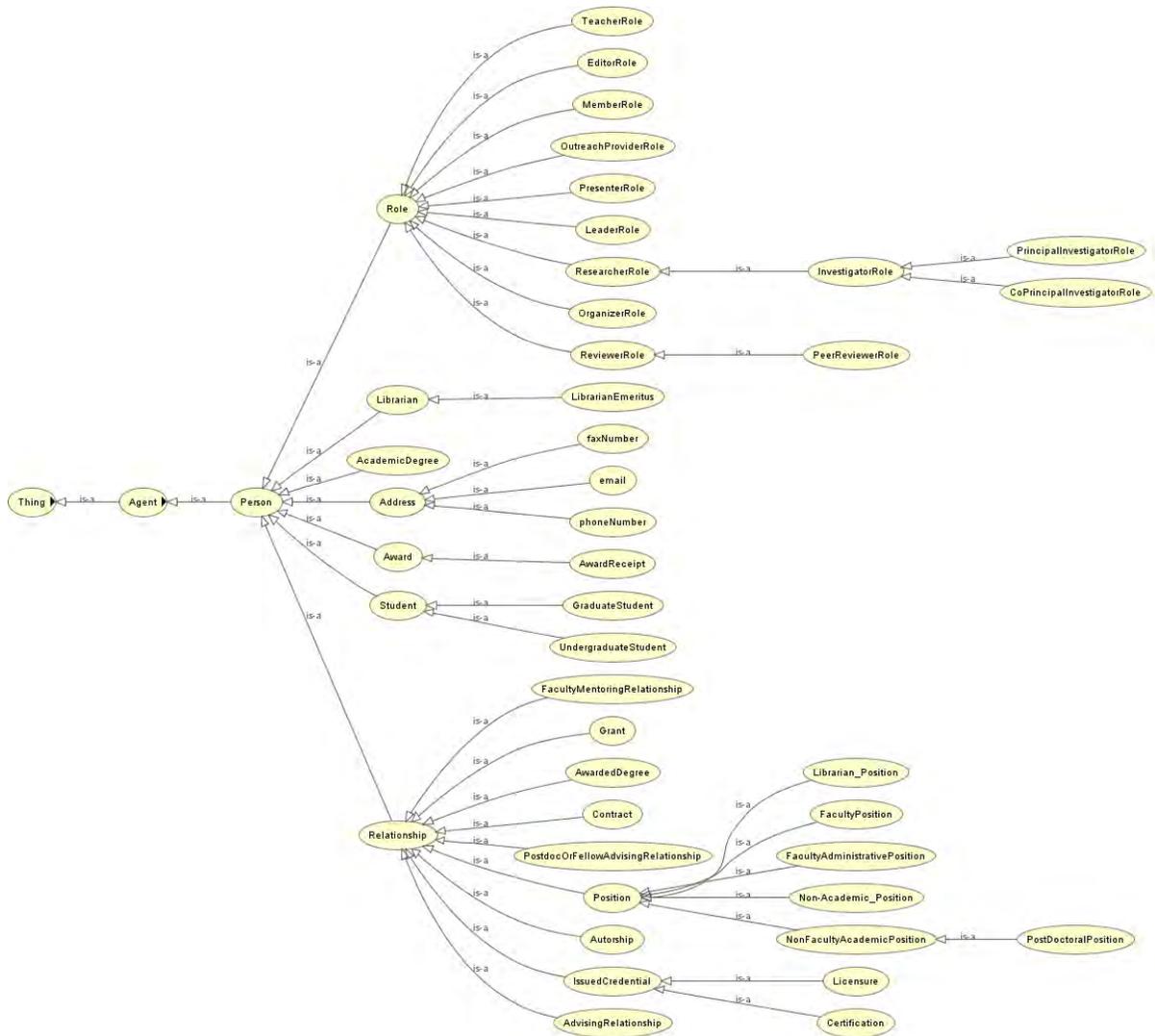


Fig. 6.48. Clase Agente y manejo de sus propiedades (elaboración propia desde la UH ontology visualizada en PROTEGE)

✓ **LOM: Learning Object Metadata**

El uso de objetos de aprendizaje ha adquirido importancia en los últimos años especialmente en las universidades. El desarrollo de iniciativas de estandarizar la manera de describirlos lo demuestra y ejemplos de esto es LOM (Learning Object Metadata). Este se refiere a la estructura de organización de los objetos de aprendizaje. Este valor puede ser especificado por varios de los términos definido en una ontología con lenguaje OWL (Soto Carrión, 2011).

En el caso de la UH ontology se ha definido dentro de la clase DOCUMENT y contiene varias especificaciones las que pueden apreciarse en la figura 6.49 que sigue a continuación. El objetivo es clasificar los tipos de recursos de aprendizaje que pudiesen gestionarse desde la ontología (Learning_Resource_Type)

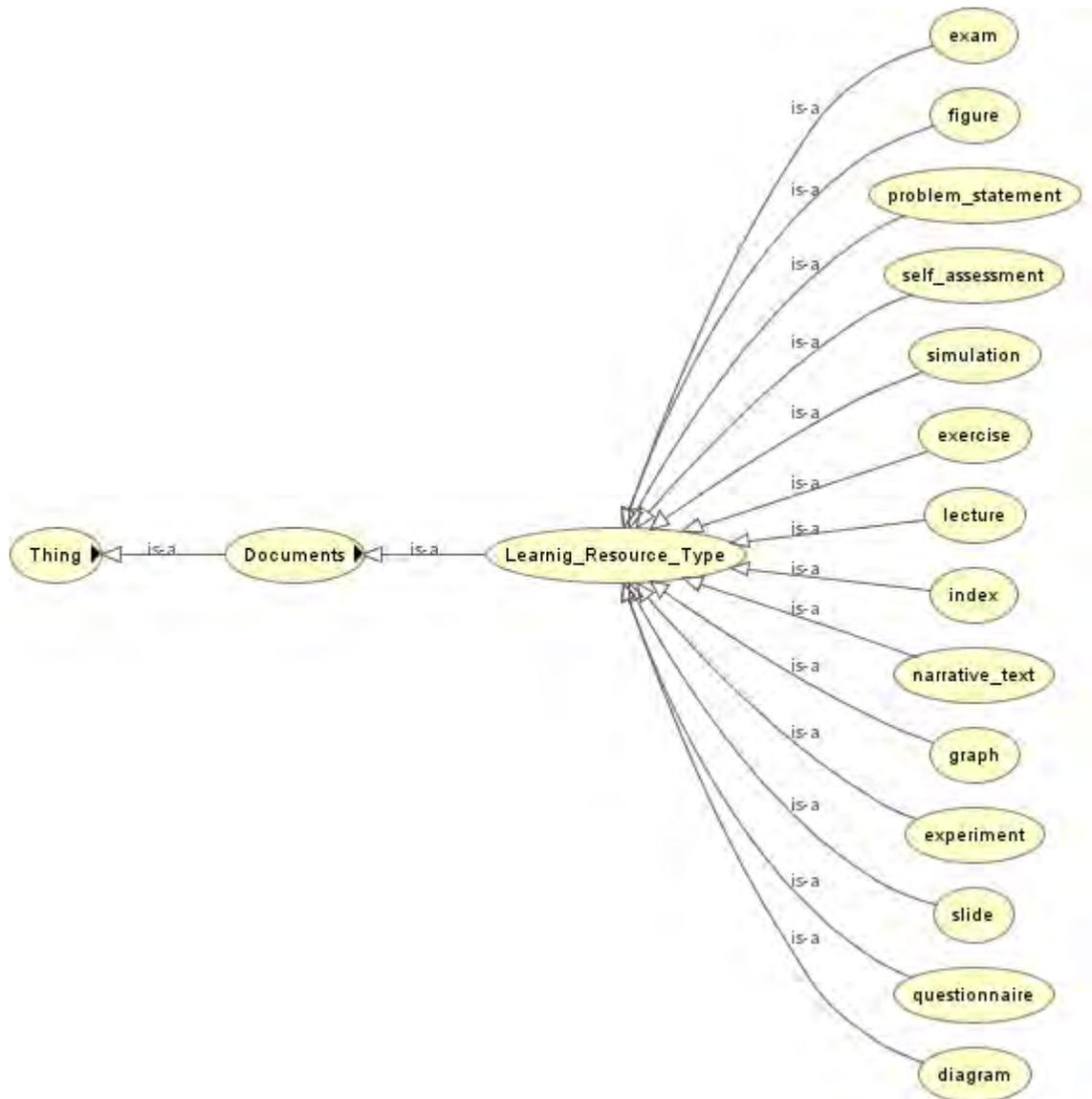


Fig. 6.49. Subclase LearningResourceType y manejo de sus propiedades (elaboración propia desde la UH ontology visualizada en PROTEGE)

6.1.6. Esquema de Anotación Semántica

La anotación semántica será el proceso mediante el cual se describan los contenidos que se asocian a un sitio web cuando se maneje el CMS UH-WEB. Generalmente, cuando la entidad que necesita el CMS tiene pocas estructuras o

dependencias en su estructura este tipo de anotación se basa en alguna de las herramientas al uso para estos fines como Annotea. Desde la óptica de esta propuesta, a consideración de la autora, la anotación semántica requiere el reconocimiento de las limitaciones del PHP 5 para conseguir descripciones de datos de forma profunda y coherente de acuerdo a las necesidades de algunas entidades. Por tanto, se determina trabajar tomando como base esquema de anotación de Annotea⁸², al que se le hacen algunas modificaciones (agregaciones).

Se proponen tres requisitos, basado en la propuesta de Navarro Galindo (2012) que deben estar presentes en el proceso de anotación: (Navarro Galindo, 2012)

- El uso de UH Ontology tanto a nivel de infraestructura durante el proceso de creación de anotaciones semánticas, como a nivel de referencia durante el proceso de asociación de significado a los textos marcados.
- El uso de algún esquema de anotación propuesto por W3C. Para este caso particular se partes de Annotea y se enriquece en función de las necesidades de la propuesta.
- El uso de estándares de la W3C para el marcado semántico. Para este caso se implementan RDF, RDFa y OWL.

6.1.6.1. Uso de Ontologías

Un CMS Semántico demanda ontologías para en su diseño, implementación y/o uso. Para este particular, la autora propone usar una ontología disponiendo de varios lenguajes ontológicos que consigan la semántica y la interoperabilidad entre las diversas fuentes de datos, recursos y formatos que se generan y gestionan en la UH y que manejará el CMS en un futuro.

⁸²Annotea es un proyecto de la W3C, que permite hacer anotaciones en las páginas Web sin que el documento original sufra ninguna transformación. Su formato principal es RDF. Tiene un estilo de anotación semiformal, donde "las anotaciones son sentencias de texto libre sobre documentos. Estas sentencias deben tener metadatos y pueden ser clasificadas de acuerdo con un esquema RDF de complejidad arbitraria definido por el usuario" NAVARRO GALINDO, J. L. 2012. *FLERSA: un sistema semántico de gestión de contenido web (S-CMS)*, Universidad de Granada.

La UH ontology dentro del modelo de anotación semántica que se proponen se encargará de lo siguiente:

- Almacenar el conocimiento generado en la UH en el cumplimiento de su función social en el ámbito docente, investigativo y extensionista y de su ejercicio administrativo. Dicha salvaguarda se realizara de manera que el CMS pueda utilizarlo en la extracción y asociación de conceptos, así como en procesos de inferencias lógicas. Así provee al CMS de suficiente semántica para que se pueda mostrar de forma eficiente la diversidad de datos que maneja
- Estructurar el conocimiento de manera que propicie la comprensión de los significados por los autómatas y por humanos en los procesos de búsqueda de información.
- Garantizar la escalabilidad del sistema utilizando formatos de metadatos aprobados por el consorcio de la W3C.
- Facilitar la búsqueda compartida y la identificación univoca de cada recurso de información.
- La reutilización de los acervos temáticos de la comunidad universitaria en forma de concepto.

No queda descartada la posibilidad de integrar al CMS en su futura implementación ontologías de dominio que favorezca el uso de un vocabulario consensuado, al aportar una taxonomía de conceptos, dada la diversidad de temas que se trabajan en la UH desde la docencia y la investigación en sus áreas.

Las ontologías tienen la finalidad de facilitar la gestión eficiente de las anotaciones por ello, en este modelo se necesita declarar como las diversas ontologías que sirven de punto de apoyo al sistema funcionarán para la realización de anotaciones especiales en el texto html. La autora coincide con el modelo de FLERSA en que el uso de múltiples lenguajes ontológicos proveerá al sistema de determinadas cuestiones entre las que se encuentran:

Facilidad para asociar y describir conceptos usando el SKOS y los etiquetas de mapeo y relación skosLabel, SkosPrefLabel, Skos:AltLabel y SkosExactMach, lo

que enriquece la propiedad Related para asociar propiedades temáticas. Por lo que con las etiquetas del SKOS es posible relacionar y controlar el vocabulario.

El vocabulario es enriquecido toda vez que se utilizan muchas ontologías cuya semántica facilita reconocer otras entidades, ya que la ontología tiene en esta aplicación la función de base de conocimiento.

Uso del RDF Schema para formar asociaciones de tipo más específico en el que se declaran materia o temáticas y unir dichas asociaciones a otras páginas web dentro del sistema con el uso de FOAF.

6.1.6.2. Annotea como Sistema de Anotación

Annotea es un software que permite el desarrollo de anotaciones compartidas y fue desarrollado ante la ineficiencia de muchas herramientas para transformar las páginas web en estructuras semánticas recuperables y representadas por un CMS. Como se señaló en epígrafe dedicado al tema en el capítulo 2, la estructura fundamental para el manejo de datos en Annotea es el RDF, aunque también maneja datos en XML y HTML.

Para la construcción teórica del modelo CMS UH-WEB, Annotea ofrece la posibilidad de anotar datos en los documentos con el manejo de la tecnología Xpioniter, sin embargo al solo circunscribirse a recursos web, resulta insuficiente para la pretendida modelación de CMS semántico para la UH, por su baja capacidad de anotación.

La mayoría de los autores reconoce que Annotea posee un esquema muy básico de anotación basado solo en elementos generales de descripción y una única clase de anotación cuyas propiedades no son suficientes para el desarrollo de anotaciones eficientes:

Annotates: Asocia una anotación con otra dentro de una página web

Author: Es el creador de la información, el responsable de la información que carga la web.

Body: Herramienta de asociación para establecer nexos entre el contenido documental y la anotación.

Context: Se refiere al contexto donde se gesta el documento y se establece como una propiedad del texto web.

Created: Expone la fecha en que se ha creado el recurso.

Modified: Declara la fecha en que se modificó el documento.

Related: Asocia al recurso principal recursos secundarios.

Las propiedades también poseen otros elementos entre los que se encuentran: Comment, seealso, question, explanation, example, advice y change.

Por tanto, puede concluirse que la propuesta de la autora utiliza Annotea con la adición de elementos que tributen la calidad del proceso de anotación semántica de los contenidos.

6.1.6.3. Infraestructura de Anotación Semántica

La anotación semántica desde el punto de vista de la web semántica se ha convertido en un proceso que permite el marcado de puntos de acceso visibles para la recuperación de la información y para la detección de entidades o datos relevantes en el texto HTML. La mayoría de las formas de anotación usadas son débiles desde el punto de vista documental. Si bien es cierto que los softwares al uso propuestos por el consorcio de la W3C para la anotación no disponen de las condiciones suficientes para el manejo de representaciones complejas de datos, tampoco puede negarse la capacidad de estos para modificarse y adoptar modelos de anotación más complejos y apegados a necesidades de múltiples clientes.

En la propuesta presentada por la autora en el modelo de CMS UH-WEB, se ha seleccionado el framework Annotea, diseñado por la W3C y también aplicado en el proyecto FLERSA (Navarro Galindo, 2012, Navarro Galindo and Samos), como base para el manejo de la estrategia de anotación. Aunque se mantiene la taxonomía de clases de infraestructura (ver Fig. 6.50), al mismo se le adicionan un conjunto de clases de anotación que hacen más potente su procesos de marcado semántico (Ver fig. 6.51 y 6.52).

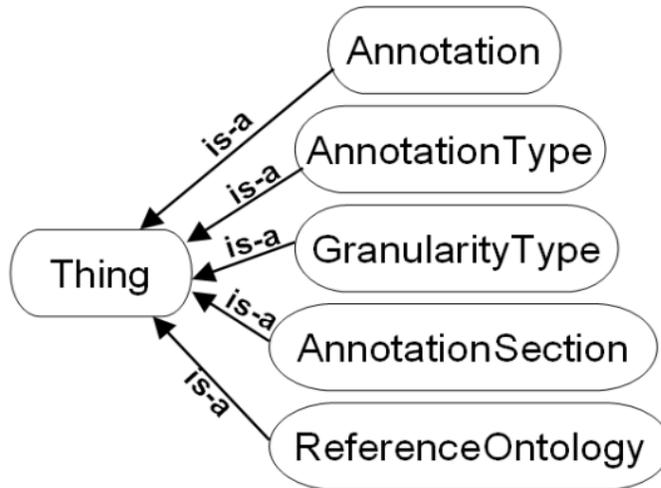


Fig. 6.50. Taxonomía de clases de infraestructura de Annotea (Navarro Galindo, 2012)

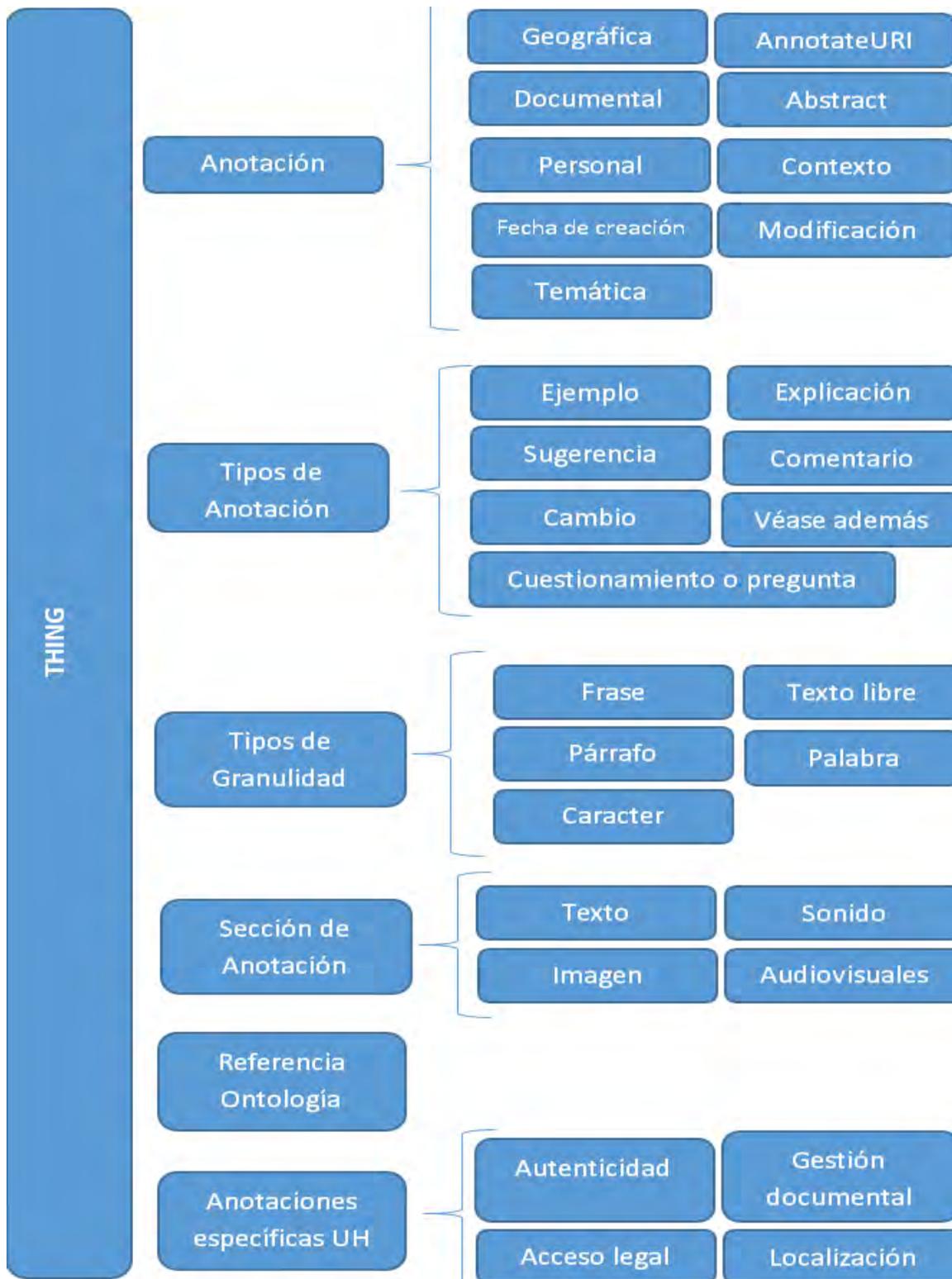


Fig. 6.51. Taxonomía de clases de infraestructura del CMS UH WEB (elaboración propia)

La ontología UH ontology se compone de varias clases de anotación basadas en diversos lenguajes ontológicos. A continuación se mostrarán las clases principales utilizadas para la anotación semántica:

Anotación ID: es la clase principal y dispone de las siguientes propiedades:

→ **Annotate URI (annotateURI):** Al igual que en Annotea y en FLERSA se dedica una anotación a asociar una anotación con la página web (recurso) sobre el que se anota. Ejemplo: <http://intranet.uh.cu/>

→ **Área Geográfica (geographic):** Se reconoce a partir de las coordenadas establecidas en la ontología.

→ **Documento (document):** Es un tipo de anotación que permite la selección de los datos bibliográficos de cada documento descrito en el sistema para ello se utilizan los lenguajes ontológicos Biba y DublinCore. Esta clase de anotación genera anotaciones de tipo de frecuencia, número, volumen, páginas, editorial.

→ **Persona (person):** Anota no solo autores de los documentos, sino toda la información asociada a la creación, modificación y manejo del documento html. Además este tipo de anotación reconoce en el texto aquellos elementos internos del texto que se asocian a personas, lo que se explicará cuando se declare en este capítulo el proceso de marcado automático que utiliza la ontología del sistema.

→ **Abstract (abstract):** Sirve para anotar el texto del documento y dentro del mismo se utiliza el sistema de marcado sugerido por Leiva para desarrollar los procesos de marcación de los elementos del texto se determina organizar las unidades textuales en segmentos (Leiva Mederos, 2011). Dichos segmentos agrupan elementos de diversos órdenes como el discursivo, el sintáctico y el comunicativo, los cuales son referentes internos de las cargas semánticas y estructurales de cada oración. Las concepciones en que se ha basado el autora para el diseño de las etiquetas obedecen a tres criterios esenciales: la secuencia en que se etiquetan los niveles en la práctica, la simbología utilizada para declarar los elementos cohesivos del texto y la presencia en el texto de elementos geográficos, personales,

documentales y de entidades. El texto se marca utilizando la ontología con sus niveles de instanciación que describe todos los procesos, maneja todas las personas de la UH.

→ **Fecha de Creación (created):** identifica mediante una etiqueta de DublinCore denominada DublinCore: Date

→ **Temática (subjct):** Identifica los temas principales que con carácter denotativo se encuentran en los documentos para ello cuenta con una el lenguaje ontológico SKOS alojado en la ontología y con las propiedades altLabel y PrefLabel, esto permite reconocer las temáticas de las que tratan los documentos mediante el análisis de texto. Generalmente las mayoría de los autores resuelven este problema usando algoritmos de agrupamiento sin explotar las bondades de la ontologías o usan las ontologías para la desambiguación de los textos, procesos de que encarece el marcado semántico, tal es el caso del algoritmo Aguirre (Aguirre, 1998) y Rigau (Rigau, 2002) el cual funciona mediante: Determinación de la totalidad de los nodos que identifican los sentidos de los vocablos a desambiguar W y los términos contextuales, lo que permite medir la densidad conceptual referente al sentido de cada término. Los elementos connotativos se marcan a partir de la propia experiencia del usuario y se aplican solo a materiales audiovisuales.

→ **Contexto (context):** Describe la posición donde se encuentra el texto o documento audiovisual que se anota.

Se establecerán además otras clases para la anotación, como puede observarse en el fig 6.51 y 6.52:

- **Tipos de anotación (type):** Asocia la anotación a un concepto dentro de la taxonomía "AnnotationType", heredada de FLERSA, indicando el tipo de granularidad dela anotación. Los tipos de anotación que se pueden realizar son: *ejemplo, sugerencia, cambio, cuestionamiento, explicación, comentario y véase además.*

- **Tipos de granularidad (granularity):** Asocia la anotación a un concepto dentro de la taxonomía "GranularityType", heredada de FLERSA, indicando el tipo de granularidad dela anotación. Los tipos de granularidad

son: *caracter, palabra, frase, párrafo o texto libre*. A sugerencia de Navarro Galindo (2012), el tipo de granularidad se debe asignar de forma automática dependiendo de las características del fragmento de texto que se anote, así como, no debe utilizarse en la anotación de textos multimedia (Navarro Galindo, 2012).

- **Identificación de las secciones de anotación (section)**: se utiliza para asociar una anotación a los formatos contenidos en un recurso y sobre el cual se realiza la anotación. Por tanto, se identifica: *texto* e *imagen* al igual que en FLRESA, se adiciona *sonido* y *audiovisuales*, dadas las características de los recursos gestionados desde la UH, especialmente los de fines educativos (learningrecourse)

- **Referencia de ontología (ontologyReference)**: establece la relación entre la anotación y las ontologías de referencia que se usan a modo de taxonomías. Se utilizará para asociar la anotación con el concepto del que se habla en ella.

- **Anotaciones específicas para la UH (annotateUH)**: Este instancia es definida para anotaciones específicas, propias y particulares de la UH que no se contemple bajo otros esquemas de anotación utilizados en la ontología. Asocia la anotación a un concepto dentro de la taxonomía "AnnotateUH", indicando el tipo de especificidad de la anotación. Para ella se definen: *autenticidad, gestión documental, acceso legal, localización*.

La anotación ha de realizarse haciendo uso del prefijo de marcado que se relaciona con el lenguaje ontológico específico (*uhvoc*), declarado para el contexto de la UH. Este se genera utilizando la siguiente URI: <https://uhvoc.cms.3/core/h>

De la misma manera que cada ontología genera prefijos de anotación que las identifica (*geo*: para anotaciones geográficas; *dc*: anotaciones DublinCore para documentos; *skos*: para las anotaciones referidas a temáticas (*skos* concept) y *vivo*: para las anotaciones referidas a la gestión de personas) para el sistema de anotaciones específicas de la UH quedan declarados los siguientes prefijos:

- o **Autenticidad:** anotaciones sobre datos de autenticidad: *uhvocDestino, uhvocConfirmacion, uhvocFirma*
- o **Gestión de documentos:** anotaciones que aportan información relativa a la gestión de derechos sobre los documentos: *uhvocPropiedad, uhvocPaternidad*
- o **Acceso legal:** anotaciones sobre restricción de uso del contenido digital y acceso legal a los documentos: *uhvocRestringidoA, uhvocVigenteHasta, uhvocClave, uhvocEncript, uhvocContrasenna,*
- o **Localización:** anotaciones referidas a la correspondencia con el sistema archivístico de gestión documental del centro: *uhvocArchCod, uhvocLocFisica.*

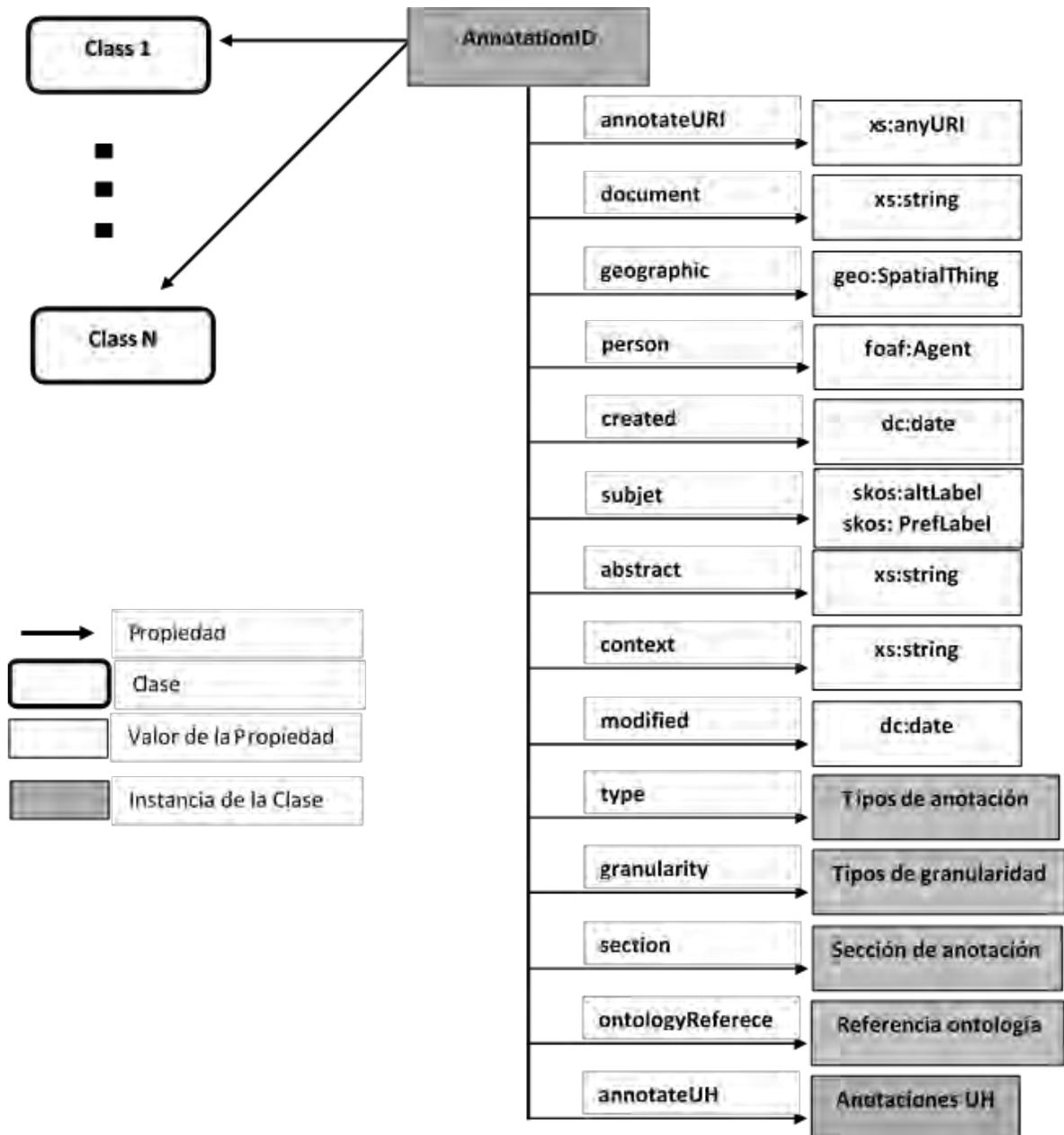


Fig. 6.52. Instancia de la clase AnnotationID (elaboración propia)

Este esquema de anotación facilitará la realización de consultas en SPARQL EndPoint.

La propuesta no contempló la anotación automática por el nivel de generalidad de las anotaciones que serían generadas y la UH se distingue precisamente por la especificidad de sus documentos, lo que requiere de

anotaciones específicas, solo obtenidas, hasta el momento, de forma manual a partir del trabajo de especialistas.

6.1.6.4. Anotación Semántica Manual

McEnery y Wilson (2001) indican que hay dos tipos básicos de anotación semántica:

- la anotación de rasgos semánticos de las palabras, como por ejemplo la anotación del sentido de cada palabra del corpus,
- la anotación de relaciones semánticas entre unidades textuales, como, por ejemplo, la anotación de roles semánticos

Para el modelo propuesto se trabaja sobre la base de la anotación semántica de palabras. Para ello se emplea la propuesta de Leiva (2011) para desarrollar los procesos de marcación de los elementos del texto a través de la organización de las unidades textuales en segmentos. Dichos segmentos aúnan elementos de diversos órdenes como el discursivo, el sintáctico y el comunicativo, los cuales son referentes internos de las cargas semánticas y estructurales de cada oración. El desarrollo de estas etiquetas permite la obtención de resúmenes extractos de un texto único, por tanto, en este apartado de la investigación nos centraremos en el diseño de las etiquetas necesarias para estructurar los textos, que serán la entrada del sistema.

A partir de ello, la autora para el diseño de etiquetas a partir de los siguientes criterios:

- la secuencia en que se etiquetan los niveles en la práctica,
- la simbología utilizada para declarar los elementos cohesivos del texto
- la presencia en el texto de elementos geográficos, personas y documentos.

Las marcas o etiquetas están trazadas en XML y están integradas por un tag de inicio y un tag de salida Ej:

- **<rdfs:comment xml:lang="en">**
The resource in which another resource is reproduced.

</rdfs:comment>

Se propone, realizar el proceso mediante la herramienta XML Marker. Es una Interfaz gráfica de usuario que permite a los anotadores tienen vistas simultáneas de todos los niveles de las anotaciones anteriores, mientras trabajaba en una tarea particular. Además, está equipado con instalaciones de comparación que permiten la inspección de inter-acuerdo anotador o rendimiento de la herramienta, expresada en precisión y recordar las medidas.

6.1.7. Recuperación de la Información

La búsqueda y recuperación de la información semántica es uno de los beneficios más impactantes de las aplicaciones semánticas que puede mostrar el CMS. Para el caso del CMS UH-WEB, el motor de búsquedas semánticas es la herramienta que recibe las consultas basadas en la UH Ontology, utilizando para ello SPARQL.

Al finalizar la implementación de UH Ontology, será capaz de dar respuestas a preguntas de competencia, que pueden servir como base para la elaboración de un test de calidad, como por ejemplo:

- *¿Cuál es el nombre del Decano de la facultad X?*
- *¿Cuál es el e-mail del administrador del sitio de la facultad X?*
- *¿Quiénes son los miembros del Consejo Científico de la facultad X?*
- *¿En qué lugar se encuentra el centro de estudio X de la Facultad Y?*
- *¿Cuál es la temática del documental X dirigido por Y?*
- *Obtener listado de participantes del evento X auspiciado por el organismo Y*
- *¿Cuáles son los cursos de postgrado que aprobados en el año X por la persona Y?*
- *Obtener las temáticas en las que investiga el profesor X*

Así, para la recuperación de informaciones específicas, como los *documentos firmados por el Decano de la Facultad de Comunicación*, se seguirían las pautas de consulta para este lenguaje (SPARQL), tal como se ejemplifica a continuación:

URL:
<http://intranet.fcom.uh.cu/>

SPARQL:

```
select ? NOMBRE where {
  ?s a bibo:DecanalFaculty.
  ?s a dc:title ?nombre.
}
```

Fig. 6.53. Consulta SPARQL para UH Ontology (elaboración propia)

Para generar las búsquedas facetadas en la ontología se recomienda la utilización de una versión modificada de la propuesta de Coppens (Coppens et al., 2009) junto a OWL2tips (Domínguez-Velasco, 2013).

OWL2tips usa SPARQL Endpoints remotos, capaces de enlazar las consultas con cada dependencia de la Universidad. Sin embargo, es importante prever que:

- Cada recurso necesitará un SPARQL Endpoint que permita construir relaciones dentro de él.
- No se puede usar una misma consulta para recuperar información en recursos diferentes, a partir del mapeo de un vocabulario.
- Es imprescindible trabajar con la especificidad en las consultas y no utilizar elementos genéricos en la descripción, por ejemplo foaf:name en vez de RDFs.
- Las consultas obligan a que los usuarios empleen lenguajes de consulta que no conocen y que les hacen complejo localizar la información, tornándolas rígidas y poco efectivas.
- Con las consultas basadas solo en SPARQL, impedirán a los usuarios recuperar toda la información que pudiesen requerir, pues hay elementos

semánticos que provienen del lenguaje OWL que quedan fuera de las consultas clásicas de SPARQL, lo que impide la elaboración de consultas más sofisticadas que exploten las posibilidades de la transitividad y la simetría cuando se trata de localizar información en varios recursos.

Por otra parte, como fortaleza, el sistema de búsqueda utiliza los elementos OWL y los enlaces de igualdad y de exclusión del lenguaje para establecer nuevas consultas dirigidas a puntos específicos de la estructura OWL, de manera que se pueda acceder a diferentes recursos de información en bases de datos remotas.

Por tanto, como solución a las dificultades previstas (anteriormente expuestas), se propone generar enlaces que procedan del lenguaje OWL y usar constructores que formen un índice de búsqueda que indexar cada elemento de las ontologías. Para esto deben almacenarse los datos en los Service Endpoint, y las propiedades de mapeo de los elementos junto a los mapas de cada clase en las ontologías.

La información se extraerá de las consultas realizadas en lenguaje SPARQL con ayuda de los constructores utilizados en la definición de la ontología, que combinados con las propiedades *OWL:sameAs*, *OWL:equivalentClass*, *OWL:equivalentProperty*, *OWL:differentFrom* y *OWL:AllDifferent* como enlaces generarán amplias capacidades para recuperar la información.

El proceso de consulta estará precoordinado para facilitar la recuperación de información, con funcionalidades preestablecidas que enlazarán los recursos. Gracias al procesador distribuido de SPARQL será posible utilizar información pre-indexada para las nuevas consultas que entren en el CMS y distribuir búsquedas derivadas de consultas generales de acuerdo con las necesidades de los usuarios potenciales del sistema.

6.2. Referencias bibliográficas

- BIZER, C. Andy Seaborne D2RQ-Treating Non-RDF Databases as Virtual RDF Graphs. *Poster at ISWC2004*.
- BRICKLEY, D. & MILLER, L. 2010. FOAF vocabulary specification 0.98, August 2010. [Online] Available: <http://xmlns.com/foaf/spec> [Accessed 10 de diciembre de 2015]
- CARABALLO-PÉREZ, Y. & RAMÍREZ-CÉSPEDES, Z. 2007. *La arquitectura de información: un análisis a partir de los procesos del Ciclo de Vida de la Información*. In V Coloquio Iberoamericano. Del papiro a la biblioteca virtual, La Habana, Cuba, 12 al 16 de marzo de 2007. [Conference paper]
- CORNELL, U. D. 2009. *VIVO:ISF* [Online]. Available: <https://wiki.duraspace.org/display/VIVO/VIVO-ISF+Ontology> [Accessed 17 de septiembre de 2015].
- D'ARCUS, B. & GIASSON, F. 2009. *Bibliographic Ontology Specification* [Online] Available: <http://biblontology.com/specification> [Accessed 27 de marzo de 2016]
- DIMIĆ SURLA, B., SEGEDINAC, M. & IVANOVIĆ, D. Year. A BIBO ontology extension for evaluation of scientific research results. *In: Proceedings of the Fifth Balkan Conference in Informatics, 2012*. ACM, 275-278.
- FLOREZ FERNANDEZ, H. A. 2013. CONSTRUCCIÓN DE ONTOLOGÍAS OWL. *Vínculos*, 4, 19-34.
- HEATH, T. & BIZER, C. 2011. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1, 1-136.
- HIDALGO-DELGADO, Y. 2015. *Marco de trabajo basado en los datos enlazados para la interoperabilidad semántica en el protocolo OAI-PMH*. Maestría, Universidad de las Ciencias Informáticas.
- HORRIDGE, M., KNUBLAUCH, H., RECTOR, A., STEVENS, R. & WROE, C. 2004. A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0. *University of Manchester*.
- ISELE, R., JENTZSCH HASSO, A., BIZER, C., VOLZ, J., PETROVSKI, P. & MANNHEIM, U. O. *Silk: The Linked Data Integration Framework* [Online]. Available: <http://silkframework.org/> [Accessed 12 diciembre 2015].

- JCORSON-RIKERT. 2011. *VIVO 1.3 Ontology overview diagram* [Online]. Universal Wiki Converter. Available: VIVO 1.3 Ontology overview diagram [Accessed 21 de agosto de 2011].
- LARIN-FONSECA, R. & GAREA-LLANO, E. 2011. Automatic Generation of the Data-Representation Ontology for Semantic Integration of Heterogeneous Geographical Data. *VII Congreso Internacional GEOMATICA*.
- LEIVA MEDEROS, A. A. 2011. *Texminer: Un Modelo para el Resumen Automático y la Desambiguación de Textos Científicos en el Dominio de Ingeniería de Puertos y Costas*. Tesis doctoral, Universidad de Granada.
- LEIVA MEDEROS, A. A., SENSO, J. A., DOMINGUEZ-VELASCO, S. & HIPOLA, P. 2013. AUTHORIS: a tool for authority control in the semantic web. *Library Hi Tech*, 31, 536-553.
- NAVARRO GALINDO, J. L. & SAMOS, J. FLERSA: Soporte a la Definición de Anotaciones y Búsquedas Semánticas en un CMS. In: *Actas de las XVI Jornadas de Ingeniería del Software y Bases de Datos*, 101. La Coruña, España.
- NAVARRO GALINDO, J. L. 2012. *FLERSA: un sistema semántico de gestión de contenido web (S-CMS)*. Tesis doctoral, Universidad de Granada.
- NONAKA, I., TAKEUCHI, H. & KOCKA, M. H. 1999. *La organización creadora de conocimiento: cómo las compañías japonesas crean la dinámica de la innovación*, Oxford University Press México DF.
- RAIMOND, Y. & ABDALLAH, S. 2007. The event ontology. [Online] Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.126&rep=rep1&type=pdf> [Accessed 16 de abril de 2016]
- RENSELAER-POLYTECHNIC-INSTITUTE & TETHERLESS-WORLD-CONSTELLATION. 2014. *Vocabularies Encoded in SKOS* [Online]. Available: <https://tw.rpi.edu/web/project/CMSPV/KeyConcepts> [Accessed 28 de noviembre de 2015].
- SOLOMOU, G. & PAPTAEODOROU, T. 2010. The use of SKOS vocabularies in digital repositories: the DSpace case. In: *Semantic Computing (ICSC), 2010 IEEE Fourth International Conference*. IEEE, 542-547.
- SOTO CARRIÓN, J. 2011. *Repositorios Semánticos de Objetos de Aprendizaje*, [Online] Available: https://books.google.com/cu/books?hl=es&lr=&id=-QBxAwAAQBAJ&oi=fnd&pg=PR7&dq=SOTO+CARRI%C3%93N,+J.+2011.+Repositorios+Sem%C3%A1nticos+de+Objetos+de+Aprendizaje,+&ots=0JnzNI902u&sig=708j78cU7k16krPmqS42NOYOFX8&redir_esc=y#v=onepage&q&f=false [Accessed 29 de abril de 2016]
- STEELE, B. 2009. *Cornell's VIVO concept will expand to connect researchers nationwide* [Online]. Available:

- www.news.cornell.edu/stories/Oct09/VIVOweb.ws.html [Accessed 7 de noviembre de 2015].
- Structured Dynamics LLC. [Online] Available: <http://bibliontology.com/> [Accessed 10 de diciembre de 2015].
- VOLZ, J., BIZER, C., GAEDKE, M. & KOBILAROV, G. 2009. Silk-A Link Discovery Framework for the Web of Data. *LDOW*, 538.
- W3C. 2012. *SKOS Simple Knowledge Organization System* [Online]. W3C. Available: <http://www.w3.org/2004/02/skos/> [Accessed 10 de diciembre de 2015].

**CAPÍTULO VII
CONCLUSIONES**

Conclusiones

Al finalizar el estudio se puede arribar a las siguientes conclusiones:

- Los Sistemas Gestores de Contenido (CMS) son herramientas informáticas estructuradas sobre la base de módulos que facilitan la creación, gestión y visualización de los contenidos, en función del producto implementado. Son sistemas dinámicos que permiten manejar de manera independiente el diseño visual del producto Web, de la gestión de contenidos. Los CMS semánticos son una tipología de CMS que aparecen en paralelo al desarrollo de la web semántica por la necesidad de dotar de significado los contenidos almacenados en páginas web tradicionales.
- El campo de desarrollo de los CMS semánticos es relativamente incipiente. Requerirán de herramientas semánticas de mayor eficiencia. Actualmente se vale de argumentos de anotación que son traslaciones de las estructuras RDF que, aunque son un indiscutible recurso para manejar semánticas y lógicas descriptivas, también son insuficientes para trasladar expresiones sintácticas dotadas de una carga de significados capaces de emular con el razonamiento humano. Por otra parte, la perspectiva teórica que los sostiene está permeada de lecturas provenientes del campo de la informática por lo que deberá articularse con mayor coherencia multidisciplinar que aporte una visión menos simplista de las anotaciones del CMS y de las estructura del mismo.
- Los elementos para describir la estructura de los CMS pueden variar de un especialista a otro. Sin embargo, para los casos estudiados en el marco de la presente investigación, la presencia de los principios que rigen el flujo de la información es lo que distingue su funcionamiento y, por tanto, la manera de estudiar los procesos, cualquiera que estos sean. Cada uno de los enfoques trata de integrar de forma sistémica, holística y sinérgica los componentes, así como la función y objetivo al que responde cada uno de manera lógica y coherente, siendo una consecuencia del funcionamiento de los CMS en la praxis.

- Los elementos estructurales de los CMS deben evolucionar con los principios de trabajo de la web semántica, haciéndose más dinámicos y enrutados a nuevos desafíos como la limpieza y la desambiguación, requisitos que han obligado a la implementación de algoritmos de aprendizaje en algunos prototipos.
- Las CI como ciencia multidisciplinar posee herramientas teóricas y prácticas para establecer nuevas rutas para el desarrollo de los CMS. La implantación del paradigma sociocognitivo sobre el paradigma físico en la descripción de información, unido a la convergencia de nuevos documentos, nuevas necesidades de información y nuevos usuarios, conllevan a esta ciencia a formular nuevas teorías en la reconcepción del esquema estructural de los CMS
- Las teorías de la CI y el uso de CMS y S-CMS, desde la concepción del producto en su dimensión cognitiva y su proceso representacional, pueden aportar las teorías necesarias para entender y enriquecer los estudios sobre la connotación semántica de los contenidos gestionados, lo que tributaría a la calidad de la información de los productos Web que con los CMS se implementan desde las CI. Este espacio ha sido poco tratado. Los conceptos de interoperabilidad, razonamiento, estándares de mapeo, se tornan indispensables hoy para hablar de un modelo de CMS coherente.
- Existen varias plataformas y herramientas de anotación para el manejo de CMS. La mayoría de estas aplicaciones se centran en el manejo de estructura a partir del PHP y emulan con la semántica pero no la consiguen **aspecto que ha sido criticado en el libro "Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data" publicado en el 2015 (Leslie, 2015)**. Las aproximaciones más importantes en el terreno de los S-CMS resuelven los problemas de la semántica usando mecanismos de anotación semántica, que se realizan de forma manual o automática. A diferencia de otros modelos (por ejemplo: Interactive Knowledge Stack (IKS) y el S-CMS FLERSA (Navarro Galindo, 2012)), la propuesta de la autora se basa en la presentación de la modelización lógica de los procesos

que deben ocurrir en el CMS semántico y sobre dicho modelo se construye su estructura de capas técnicas.

- Los CMS de la UH reportan escaso valor en desarrollo de la gestión de información en la universidad. Desde el punto de vista de técnico están sustentados en Joomla o Drupal. Joomla posee una arquitectura compleja y rígida, deficiente para mantener una estructura de datos en RDF, proceso que se complejiza cuando se intenta modificar el código de este usando lenguaje de programación PHP. Drupal tiene un soporte RDF incompleto y con escaso desarrollo, por lo que muchos desarrolladores para lograr semántica, interoperabilidad y datos enlazados prefieren ir a aplicaciones basadas en SILK, DR2K o Pubby. Cuando se evalúa la calidad de los datos gestionados por estos CMS se puede observar que los parámetros son bajos muchos se deben a duplicidades y a la falta de significado generado en consulta poco eficientes.
- El CMS UH-Web es una herramienta que, aunque aún se maneja solo en el orden teórico, constituye un aporte novedoso como diseño de un CMS. Su principal fortaleza recae en el diseño de una capa semántica dotada de razonamiento, búsquedas en SPARQL, almacén de tripletas RDF y datos enlazados, cualidades que hacen que el sistema sean muy superior a los modelos existentes para el manejo de la semántica. A estas propiedades semánticas se une el uso extensivo de ontologías de diversa índole que aportan mayor semántica al sistema de anotación que ha sido establecido para describir y diferenciar los documentos electrónicos. La propuesta de la ontología para homogenizar y gestionar los datos online de la UH promete significativos avances en los procesos sustantivos de gestión de la información institucional, tanto de sus recursos de información, datos personales y organizacionales. El modelo personalizado para la UH tributa a la construcción de sitios personalizados, con mejor organización de los contenidos gestionados en la universidad y por tanto, incide directamente en los procesos de búsqueda y recuperación de la información de las intranets y sitios construidos.

- UH Ontology ofrecerá al CMS UH-WEB la posibilidad de normalizar los atributos de los metadatos aplicables a los documentos; crear una red de relaciones que aporte especificación y fiabilidad; la reutilización del conocimiento existente; compartir conocimiento en pos de hacer explícito los diferentes criterios, así como la integración de diferentes perspectivas de los usuarios, fomentando el trabajo cooperativo al funcionar como soporte común de conocimiento entre las distintas áreas y estructuras de la UH. Como consecuencia de todo ello, la interoperabilidad entre los distintos sistemas existentes en la UH y el establecimiento de modelos normativos, contribuirán a la creación de la semántica de un sistema y un modelo para poder extenderlo y transformarlo en los diferentes contextos de la Universidad.

Referencias bibliográficas

- LESLIE, F. 2015. *Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data*, Apress. [Online] Available: https://books.google.com.cu/books?hl=es&lr=&id=mGYnCgAAQBAJ&oi=fnd&pg=PP3&dq=LESLIE,+F.+2015.+Mastering+Structured+Data+on+the+Semantic+Web:+From+HTML5+Microdata+to+Linked+Open+Data,+Apress.+&ots=x5gIN4mws0&sig=hXDbhE8PB85V0QU9Y0xpeBYWmc&redir_esc=y#v=onepage&q=LESLIE%2C%20F.%202015.%20Mastering%20Structured%20Data%20on%20the%20Semantic%20Web%3A%20From%20HTML5%20Microdata%20to%20Linked%20Open%20Data%2C%20Apress.&f=false [Accessed 8 de abril de 2016]
- NAVARRO GALINDO, J. L. 2012. *FLERSA: un sistema semántico de gestión de contenido web (S-CMS)*. Tesis doctoral, Universidad de Granada.

CAPÍTULO VIII
TRABAJO FUTURO

Trabajo Futuro

En este apartado se presentan algunas líneas que conducirán el desarrollo del modelo desde una perspectiva de futuro.

✓ Interoperabilidad y Datos enlazados

El desarrollo de la interoperabilidad es una de las destrezas a la que se encamina el desarrollo de los CMS semánticos. Aplicaciones a nivel mundial exponen sus datos en formato OAI-PMH o OAI-ORE, lo que conduce al manejo eficiente de datos externos. Para lograr la interoperabilidad hay que conseguir que los CMS puedan manejar, desde su soporte nativo, protocolos de comunicación TCP/IP, intercambiar datos mediante formatos de intercambio como el XML.

La mayoría de los CMS tienen dificultades para publicar metadatos siguiendo los principios de los datos enlazados. Actualmente se desarrollan estándares y tecnologías en el contexto de la web semántica para garantizar la interoperabilidad entre los sistemas de información que no podrían manejarse dentro de los CMS y conllevarían transformaciones en los mismos. Los esfuerzos en el terreno muestran soluciones incapaces de transformar metadatos simples en grafos de tipo RDF y aplicadas a la actividad documental. Algunas herramientas, entre las que se encuentran Liferay y Drupal, poseen SPARQL Endpoints capaces de manejar datos enlazados y conseguir interoperabilidad, pero las aspiraciones para el amplio manejo, son remotas y actualmente están poco desarrolladas.

Entre los trabajos más sólidos en este sentido puede referirse:

- Unión de Catálogos de Suecia (LIBRIS) (Malmsten, 2008). Para la transformación de los registros bibliográficos en formato MARC21, donde se implementó un RDF Server Wrapper. Este servidor utiliza una aproximación basada en eXtensible Stylesheet Language Transformations (XSLT) para realizar dicha tarea.

- OAI2LOD Server (Haslhofer and Isaac, 2011, Haslhofer and Schandl, 2010, Haslhofer and Schandl, 2008) es una herramienta para la transformación de metadatos bibliográficos diseminados mediante OAI-PMH a datos enlazados, aunque la misma posee limitaciones de tipo estructural (Hidalgo-Delgado, 2015, Coppens et al., 2009).
- Marimba (Rodriguez Greenberg et al., 2013, Vila-Suero et al., 2013) es de las soluciones más complejas, pero de alto costo económico. Marimba es una herramienta para la transformación de registros en MARC21 a grafos RDF. La herramienta según Hidalgo (2015) proporciona un entorno para la alineación de clases y propiedades de ontologías con los elementos de metadatos del formato MARC21 (Hidalgo-Delgado, 2015).
- XMLRDF (De Boer et al., 2012) se describe como una herramienta utilizada para transformar registros de metadatos bibliográficos en grafos RDF. La transformación es realizada en dos etapas. Según Hidalgo (Hidalgo-Delgado, 2015), en la primera etapa se realiza una conversión sintáctica de los archivos XML fuentes en grafos RDF crudos. El RDF obtenido debe ser lo más cercano posible al metadato original. En una segunda etapa el RDF obtenido de esta manera es reestructurado y enriquecido. La herramienta proporciona una biblioteca dinámicamente extendida de rutinas en Prolog para tareas comunes de conversión.
- OAI-PMH RDFizer fue gestada en 2007 por el proyecto SMILE del Massachusetts Institute of Technology (MIT), por Stefano Mazzocchi (Hidalgo-Delgado, 2015). Esta herramienta convierte metadatos diseminados mediante el protocolo OAI-PMH en grafos RDF. La herramienta utiliza hojas de estilo XSLT para transformar datos. Indiscutiblemente todos los sistemas poseen limitaciones para generar datos enlazados en un CMS, primero la entrada de los datos se sustenta únicamente en una fuente de datos por tanto no permitiría manejar datos heterogéneos dentro del CMS, ya que toman como entrada una única fuente de datos, usan transformaciones a partir de XSLT, lo que hace que cuando se genere un cambio en la fuente de datos en el servidor haya que establecer cambios en las hojas de estilo, y editarlas manualmente, además los metadatos

siempre no tiene la calidad requerida en la ontologías, lo que obliga antes de interoperar sea vital hacer tareas de preprocesamiento de datos antes de que el CMS publique la información.

✓ **Recuperación de Información**

Se espera un proceso evolutivo continuo en el campo de la recuperación de la información que favorezca la aplicación y desarrollo del modelo de UH Web, así como mejores técnicas de inteligencia artificial.

Los adelantos más concretos en materia de desarrollo de los procesos y herramientas de búsqueda y recuperación de la información, se materializan en la capacidad de los buscadores para interrogar gran cantidad de datos y estructurarlos de forma inteligente. No obstante, la ausencia de agentes que manejen de forma coherente este propósito, hacen revisar los núcleos de razonamiento de las aplicaciones constantemente. Se necesita que la inteligencia artificial trabaje sobre mejores algoritmos de aprendizaje, que han de trascender la distribución de reglas y reutilización de la información de otras páginas web, aspectos que no han sido desarrollados en su totalidad.

La búsqueda en un CMS semántico se basa en herramientas de consulta sustentadas en SPARQL. La capacidad de razonamiento de estos aún es baja, pues se aspira a que emulen con el pensamiento humano, y en este afán solo son capaces de calcar algunas reglas de manejo de datos, pues SPARQL aún es poco operativo para generar búsquedas totalmente eficientes.

Actualmente, el porcentaje de las páginas web que han adoptado el RDF es bajo y se mantienen en un formato XHTML y HTML5. Esto obliga pensar sobre la forma en que se manejan los datos.

✓ **Marcado Semántico como herramienta para la recuperación de la información**

Es importante mejorar las técnicas de marcado semántico para implementar UH Web, la mayoría de las que existen están descritas para datos factuales o metadatos documentales y escasamente desarrolladas para la marcación automática de atributos extendidos a lo largo de un texto. Los mecanismos con que se ejecuta el marcado demandan una renovación modélica que explote técnicas de aprendizaje semi-supervizado, buscando complementar los procesos mentales que se realizan en el marcado manual.

Las técnicas de marcado semántico que utilizan microformatos, *RDFa* y microdatos, se considera que aún están alejadas de las necesidades de riqueza semántica, factor indispensable para lograr una eficiente recuperación de la información:

- Microformatos es una forma de marcado muy utilizada con el XHTML (versiones 1 y 5) y HTML (versiones 4 y 5) y se combina en la actualidad con CCS. Consiste en manejar el atributo `class` en la detección de propiedades, clases y vocabularios de forma combinada con CSS. La principal deficiencia que tiene alta carga de ambigüedad, surgida al **combinar en un mismo atributo "class"** y otros formatos.
- RDFa gestionada por el consorcio de la W3C, esta es compatible con los formatos de la web semántica. RDFa puede usarse con XHTML (v 1 y 5), HTML (4 y 5), además con XML, SVG (*Scalable Vector Graphics*), ePUB y OpenDocument, pero su uso es tan complejo que ha obligado al replanteamiento de nuevas clases y a la aparición de nuevos estándares homólogos como RDFaLite.
- Microdatos: Es una forma de marcado para el HTML 5 coordinados por W3C. Se utiliza en con dos serializaciones definiendo una DTD con HTML 5 y otra similar con XHTML 5.

Otro elemento a considerar en el intercambio de datos es JSON (acrónimo de Java Script Object Notation). Es un subconjunto de la notación literal de objetos de Java Script. Su formato es útil para la serialización de datos y mensajería. Como lenguaje, es más sencillo que XML pero, hasta el momento, JSON se emplea habitualmente en entornos donde la fuente de

datos es completamente fiable y no se concede mucha relevancia al procesamiento XSLT para la gestión de los datos en el cliente. No obstante en los últimos años es frecuente el uso de JSON y XML en la misma aplicación. Por ejemplo, una aplicación de cliente que integra datos de Google Maps con datos meteorológicos en SOAP hacen necesario soportar ambos formatos.

En 2014 el w3c recomienda una serialización basada en JSON para Linked Data, JSON-LD. Su sintaxis está diseñada para integrarse fácilmente en los sistemas implementados que ya usan JSON, y proporciona un camino de actualización sin problemas de JSON a JSON-LD. Está pensado principalmente para ser una manera de utilizar datos vinculados en entornos de programación basados en la Web, para construir servicios Web interoperables, y almacenar datos vinculados en los motores de almacenamiento basadas en JSON.

JSON-LD se puede utilizar para expresar datos semánticos marcados en otros formatos de datos vinculados, como RDFa, microformatos y microdatos, por lo que puede asumirse como una alternativa natural al uso de dichos formatos.

Las herramientas desarrolladas para el marcado en los CMS, aunque los dotan de algunas capacidades para el desarrollo del proceso, no están en total definición y son insuficientes para lo que se necesita. Por ejemplo:

- Drupal, tiene el módulo Semantic markup editor, OpenCalais y Alchemy, que utilizan APIs para el marcado automático de los contenidos.
- MediaWiki, dispone de las herramientas Semantic MediaWiki Síndice y OpenCalais, con servicios que facilitan el manejo de la información y el marcado automático de atributos documentales.

✓ **Integración de Herramientas de Publicación de Datos**

Las transformaciones en los métodos de publicación de datos están obligando a manejar nuevos lenguajes de declaración para resolver los problemas de

alineación de datos. Esta capacidad para el tratamiento no las tienen los CMS actuales, pero existen herramientas capaces de facilitar esto y, al ser software open source, permiten integrarse con nuevos desarrollos.

Para llevar a cabo UH Web se necesita integrar en su estructura herramientas disímiles:

- El uso de D2RO (Bizer, 2004) en la capa semántica de un CMS facilitaría el tratamiento de las bases de datos relacionales como grafos RDF y el acceso a estos mediante SPARQL.
- Con Silk (Volz et al., 2009, Isele et al.) sería posible generar enlaces entre elementos de datos basados en la especificación de enlazado proporcionada por el usuario.
- Con Pubby (Cyganiak and Bizer, 2008): un sistema de datos enlazados que proporciona vistas HTML sobre los recursos existentes en un grafo RDF almacenado en un almacén de tripletas es posible generar nuevas visualizaciones e integrar más los datos si estos se manejan en una plataforma.

✓ **Transformaciones modélicas, de formación y normativas de construcción:**

Para construir UH Web se necesita desmontar los modelos clásicos de desarrollo de CMS que si bien tienen un ciclo de vida muy argumentado y probado nada aportan en el terreno de la web semántica, pues encapsulan el manejo técnico de los datos a las etapas primarias de estos y lastran la representación de los datos. Lo más importante es ofrecer, representaciones que expliciten y compartan, no que excluyan teorías o enfoques, que se inserten en la dinámica del razonamiento contextual y así reconozcan en la práctica los basamentos del enfoque sociocognitivo, familiarizados por razones de educación y de cercanía generacional con la integración tecnológica.

La formación para el cambio es vital, la perspectiva transdisciplinar y multidisciplinar es compleja, por lo que se demandan nuevas concepciones en aspectos tan disímiles como:

- Visualización de información
- Recuperación de Información
- Manejo de Datos
- Técnica de Limpieza de datos que exploten heurísticas complejas.

Para concebir un modelo de CMS se demanda de una academia sujeta y consciente del cambio, que no sostenga sus pilares en el paradigma exclusivamente humanista y que empiece a generar profesionales imbricados de los procesos de las tecnologías de la información.

Hay que observar las mejores prácticas en manejo de datos, lo que conllevará al desarrollo de políticas de construcción de datasets y manejo de nuevas regulaciones para transmitir y consumir datos.

Referencias bibliográficas

- COPPENS, S., MANNENS, E. & VAN DE WALLE, R. 2009. Disseminating heritage records as linked open data. *International Journal of Virtual Reality*, 8, 39.
- CYGANIAK, R. & BIZER, C. 2008. *Pubby-a linked data frontend for sparql endpoints* [Online]. Available: <http://wifo5-03.informatik.uni-mannheim.de/pubby/> [Accessed 17 enero 2016].
- HASLHOFER, B. & ISAAC, A. 2011. The Europeana Linked Open Data Pilot. In: *International Conference on Dublin Core and Metadata Applications*, The Hague, The Netherlands, 94 - 104.
- HASLHOFER, B. & SCHANDL, B. 2008. The OAI2LOD Server: Exposing OAI-PMH metadata as linked data. In: *International Workshop on Linked Data on the Web (LDOW2008)*, co-located with WWW 2008, Beijing, China.
- HASLHOFER, B. & SCHANDL, B. 2010. Interweaving OAI-PMH data sources with the linked data cloud. *International Journal of Metadata, Semantics and Ontologies*, 5, 17-31.
- HIDALGO-DELGADO, Y. 2015. *Marco de trabajo basado en los datos enlazados para la interoperabilidad semántica en el protocolo OAI-PMH*. Maestría, Universidad de las Ciencias Informáticas.
- ISELE, R., JENTZSCH HASSO, A., BIZER, C., VOLZ, J., PETROVSKI, P. & MANNHEIM, U. O. *Silk: The Linked Data Integration Framework* [Online]. Available: <http://silkframework.org/> [Accessed 12 diciembre 2015].
- LANTHALER, M., & GÜTL, C. (2012). On using JSON-LD to create evolvable RESTful services. In: *Proceedings of the Third International Workshop on RESTful Design*, 25-32.
- MALMSTEN, M. 2008. Making a library catalogue part of the semantic web. *Metadata for Semantic and Social Applications: Proceedings of the International Conference on Dublin Core and Metadata Applications: Berlin, 22-26 September 2008: DC 2008*: Berlin, Germany: Universitätsverlag Göttingen.
- RODRIGUEZ GREENBERG, E. M., GEMA BUENO DE LA FUENTE, J., VILA-SUERO, D. & GÓMEZ-PÉREZ, A. 2013. datos. bne. es and MARiMbA: an insight into library linked data. *Library Hi Tech*, 31, 575-601.
- SPORNY, M., KELLOGG, G., LANTHALER, M., & W3C RDF Working Group. (2014). JSON-LD 1.0: a JSON-based serialization for linked data. W3C Recommendation, 16.

- VILA-SUERO, D., VILLAZÓN-TERRAZAS, B. & GÓMEZ-PÉREZ, A. 2013. datos. bne. es: A library linked dataset. *Semantic Web*, 4, 307-313
- VOLZ, J., BIZER, C., GAEDKE, M. & KOBILAROV, G. 2009. Silk-A Link Discovery Framework for the Web of Data. *LDOW*, 538.

BIBLIOGRAFÍA CONSULTADA

Bibliografía Consultada

- AGUILERA, F. 2016. Entrevista a especialista de la Dirección de Formatación [27 de enero de 2016]. Universidad de La Habana. La Habana, Cuba
- AINOZA FARLED, J. 2007. Estudio y evaluación de sistemas gestores de contenidos web. Tesis de maestría, Universitat Politècnica de Catalunya. [Online]. Available: <http://hdl.handle.net/2099.1/5276> [Accessed 18 de enero de 2016].
- ALBUQUERQUE, A. 1998. Los tres mundos de la Ciencia de la Información. *Ciencias de la Información*, 29 (3).
- ALOR-HERNÁNDEZ, G., SÁNCHEZ-RAMÍREZ, C., CORTES-ROBLES, G., RODRÍGUEZ-GONZÁLEZ, A., GARCÍA-ALCARAZ, J. L. & CEDILLO-CAMPOS, M. G. 2014. BROSEMWEB: A brokerage service for e-Procurement using Semantic Web Technologies. *Computers in Industry*, 65, 828-840.
- ALVITE DÍEZ, M. L. 2012. El uso de vocabularios controlados en los sistemas de información jurídica: evolución y tendencias actuales de representación. *Scire*. 18 (1), 29- 39.
- ANTONIOU, G., FRANCONI, E. & VAN HARMELEN, F. 2005. Introduction to semantic web ontology languages. In: *Reasoning web*, 1-21. Springer Berlin Heidelberg.
- ARENCIBIA JORGE, R. 2010. *La colaboración científica entre Cuba y España en el Web of Science 1999-2003*. Tesis doctoral, Universidad de Granada.
- ARPIREZ, J., GÓMEZ-PÉREZ, A., LOZANO, A. & PINTO, H. S. 1998. 2Agent: An ontology-based WWW broker to select ontologies. In: *Workshop on Applications of Ontologies and PSMs*, Brighton, England.
- ASTESIANO, E., BIDOIT, M., KIRCHNER, H., KRIEG-BRÜCKNER, B., MOSESSE, P. D., SANNELLA, D. & TARLECKI, A. 2002. CASL: the common algebraic specification language. *Theoretical Computer Science*, 286, 153-196.
- BAEZA-YATES, R. & RIBEIRO-NETO, B. 1999. *Modern information retrieval*, ACM press New York.
- BAO, J. & MOTIK, B. 2009. rdf: PlainLiteral: A Datatype for RDF Plain Literals. *World Wide Web Consortium (W3C) Recommendation 27 October 2009*. [Online] Available: <https://www.w3.org/2007/OWL/draft/ED-rdf-plain-literal-20090917/all.pdf> [Accessed 8 de abril de 2016]
- BAO, J., KENDALL, E. F., MCGUINNESS, D. L., PATEL-SCHNEIDER, P. F., DING, L. & KHANDELWAL, A. 2009. Owl 2 web ontology language quick reference guide. *World Wide Web Consortium (W3C) Recommendation 27 October 2009*. [Online] Available: <https://www.w3.org/TR/2009/WD-owl2-quick-reference-20090421/all.pdf> [Accessed 8 de abril de 2016]

- BECERRA OCAMPO, O. F. 2013. Clase 3- La semántica y la Lingüística. Lenguaje y comunicación [Online]. Available: http://lenguajeycomprension.blogspot.com/2013_01_01_archive.html [Accessed 5 de abril de 2016].
- BECHHOFER, S. 2009. OWL: Web ontology language. In: *Encyclopedia of Database Systems*. Springer US. [Online] Available: http://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_1073 [Accessed 8 de abril de 2016]
- BECKETT, D. 2014. RDF 1.1 N-Triples. *World Wide Web Consortium, Recommendation REC-n-triples-20140225*. [Online] Available: <http://www.w3.org/TR/2014/REC-n-triples-20140225/> [Accessed 7 de abril de 2016]
- BECKETT, D., BERNERS-LEE, T., PRUD'HOMMEAUX, E. & CAROTHERS, G. 2014. RDF 1.1 Turtle. *World Wide Web Consortium*. [Online] Available: <https://www.w3.org/TR/2014/REC-turtle-20140225/> [Accessed 7 de abril de 2016]
- BEGOÑA RODRÍGUEZ, M., BAZÁN, P. & HUNGARO, A. M. 2008. *Espacios virtuales con valor educativo. Espacios virtuales con valor educativo* [Online]. Available: http://www.linti.unlp.edu.ar/uploads/docs/espacios_virtuales_con_valor_educativo.pdf [Accessed 24 de marzo de 2013].
- BELTRAN-GÓMEZ, O. 2013. Implementación de algoritmos para la construcción de proveedores de datos OAI. *educonector. info*, 68.
- BENITO-OSORIO, D., PERIS-ORTIZ, M., ARMENGOT, C. R. & COLINO, A. 2013. Web 5.0: the future of emotional competences in higher education. *Global Business Perspectives*, 1, 274-287.
- BERNERS-LEE, T. 2000. *Semantic Web on XML*. In: *Slides from XML 2000 conference*, Washington DC, USA. [Online] Available: <https://www.w3.org/2000/Talks/1206-xml2k-tbl/slide1-0.html> [Accessed 19 de septiembre de 2015]
- BERNERS-LEE, T., HENDLER, J. & LASSILA, O. 2001. The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284, 34-43.
- Biblioteca Nacional de Australia. 2003. Directrices para la preservación del Patrimonio Biblioteca Nacional de Australia: División de la Sociedad de la Información. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
- BIDOIT, M. & MOSSES, P. 2004. CASL User Manual, Introduction to Using the Common Algebraic Specification Language. LNCS 2900. Springer-Verlag.
- BIZER, C. & CYGANIAK, R. 2014. RDF 1.1 TriG. *W3C recommendation*, 110. [Online] Available: <http://www.w3.org/TR/trig/> [Accessed 7 de abril de 2016]

- BIZER, C. & SCHULTZ, A. 2009. The Berlin SPARQL Benchmark. *International Journal on Semantic Web and Information Systems*, 5, 1-24.
- BIZER, C. 2004. Andy Seaborne D2RQ-Treating Non-RDF Databases as Virtual RDF Graphs. *Poster at ISWC2004*.
- BLÁZQUEZ, J., FERNÁNDEZ, M., GARCÍA-PINAR, J. M. & GÓMEZ-PÉREZ, A. 1998. Building ontologies at the knowledge level using the ontology design environment. In: *Proceedings of the Eleventh Workshop on Knowledge Acquisition, Modeling and Management, KAW'98*, Banff, Canada.
- BOIKO, B. 2001. Understanding content management. *Bulletin of the American Society for Information Science and Technology*, 28, 8-13.
- BOLDÚ 2011. Curso: "Sistemas de gestión de contenidos Digitales (SGCD): Tema 4. TIPUS DE CMS". En: Màster en Gestió de Continguts Digitals (curso 2010-2011). Universidad de Barcelona, España: [notas de clase].**
- BOSCH, M. 2001. Documentos y lenguaje de marcado: conceptos, problemas y tendencias. *El Profesional de la Información*, 10, 4-9.
- BOWMAN, C. M., DANZIG, P. B., HARDY, D. R., MANBER, U. & SCHWARTZ, M. F. 1994. Harvest: A scalable, customizable discovery and access system. DTIC Document. (No. CU-CS-732-94). COLORADO UNIV AT BOULDER DEPT OF COMPUTER SCIENCE.
- BRAMSCHER, P. F. & BUTLER, J. T. 2006. LibData to LibCMS - One library's evolutionary pathway to a content management system. *Library Hi Tech*, 24, 14-28.
- BRATSAS, C., BAMIDIS, P., DIMOU, A., ANTONIOU, I. & IOANNIDIS, L. 2012. Semantic CMS and wikis as platforms for Linked Learning. In: *2nd Int. Workshop on Learning and Education with the Web of Data-24th Int. World Wide Web Conference*.
- BRICKLEY, D. & GUHA, R. V. 2014. RDF Schema 1.1. *W3C recommendation*, 25, 2004-2014. [Online] Available: <http://www.w3.org/TR/rdf-schema> [Accessed 19 de septiembre de 2015]
- BRICKLEY, D. & MILLER, L. 2010. FOAF vocabulary specification 0.98, August 2010. [Online] Available: <http://xmlns.com/foaf/spec> [Accessed 10 de diciembre de 2015]
- BRICKLEY, D. & MILLER, L. 2014. FOAF Vocabulary Specification 0.99. Namespace Document 14 January 2014-Paddington Edition. [Online] Available: <http://xmlns.com/foaf/spec/> [Accessed 7 de abril de 2016]
- CALDÓN, E. F., URIBE, G., LÓPEZ, D. M., DE OLIVEIRA, J. P. M. & WIVES, L. K. 2010. Mecanismos de anotación semántica de Contenidos en Plataformas de Redes Sociales. *Cuadernos de Informática*, 5, 89-99.
- CARABALLO-PÉREZ, Y. & RAMÍREZ-CÉSPEDAS, Z. 2007. *La arquitectura de información: un análisis a partir de los procesos del Ciclo de Vida de la*

- Información.** In V Coloquio Iberoamericano. Del papiro a la biblioteca virtual, La Habana, Cuba, 12 al 16 de marzo de 2007. [Conference paper]
- CAROTHERS, G. 2014. Rdf 1.1 n-quads. *W3C recommendation, Feb.* [Online] Available: <http://www.w3.org/TR/n-quads> [Accessed 19 de septiembre de 2015]
- CARRILLO POZAS, A. 2014. Ontologías: Definición, metodologías y buenas prácticas para su construcción. In: *Conferencia impartida en la Facultad de Ciencias de la Documentación de la UCM* (6 de mayo de 2014), Biblioteca Nacional de España, España. [Online] Available: <http://es.slideshare.net/bne/ontologas-definicion-metodologas-y-buenas-prcticas> [Accessed 13 de enero de 2016]
- CASTELLS, M. 2002a. La dimensión cultural de Internet. Cultura XXI. [Online] Available: <http://www.uoc.edu/culturaxxi/esp/articulos/castells0502/castells0502.html> [Accessed 30 de septiembre de 2013].
- CASTELLS, P. 2002b. Aplicación de técnicas de la web semántica. In: *Workshop de investigación en entornos de interacción colectiva (COLINE'02)*. Granada, España. [Online] Available: http://moodle2.unid.edu.mx/dts_cursos_md/pos/TI/BE/AM/12/tecnicas_de_la_web_semantica.pdf [Accessed 23 de abril de 2016]
- CASTRO FERNÁNDEZ, R. 2008. Representación del Conocimiento. Web Semántica. *Universidad Carlos III de Madrid*. [Online] Available: <http://www.it.uc3m.es/jvillena/irc/practicas/08-09/05.pdf> [Accessed 19 de septiembre de 2015]
- CAVIERES ABARCA, A., FREDER MENA, S. & RAMÍREZ NOVOA, A. 2010. Tesoros y Web Semántica: Diseño metodológico para estructurar contenidos Web mediante SKOS-Core. *Serie bibliotecología y gestión de información*, 1-64.
- CENTRO DE ESTUDIOS DE ADMINISTRACIÓN PÚBLICA. 2015. Quiénes somos [Online. Available: <http://www.ceap.uh.cu/> [Accessed 13 de enero de 2016].
- CENTRO DE ESTUDIOS DE LA ECONOMÍA CUBANA DE LA UNIVERSIDAD DE LA HABANA. 2016. ¿Quiénes somos? [Online]. Available: http://www.ceec.uh.cu/que_es_el_ceec [Accessed 13 de enero de 2016].
- CENTRO DE ESTUDIOS DEMOGRÁFICOS DE LA UNIVERSIDAD DE LA HABANA. 2012. Página de inicio de la intranet [Online]. Available: <http://intranet.cedem.uh.cu/> [Accessed 15 de enero de 2016].
- CENTRO DE ESTUDIOS PARA EL PERFECCIONAMIENTO DE LA EDUCACIÓN SUPERIOR (CEPES) DE LA UNIVERSIDAD DE LA HABANA. 2012. ¿Quiénes somos? [Online]. La Habana. Available: <http://cepes.uh.cu> [Accessed 13 de enero de 2016].

- CENTRO DE INVESTIGACIONES MARINAS. 2015. Informe de Autoevaluación Institucional (2010 - Junio 2015).
- CHIARO, N., GARAT, D., DAMONTE, P. & ACCUOSTO, P. 2005. Generación semiautomática de una ontología para una red de ONG. In: **XI Congreso Argentino de Ciencias de la Computación**. Argentina.
- CHOWDHURY, G. 2010. *Introduction to modern information retrieval*, Facet publishing.
- CHRIST, F. 2012. Interactive Knowledge Stack: A Software Architecture for Semantic Content Management Systems. In: PADERBORN, U. (ed.) **Architekturen 2012**. Paderborn.
- CMS-MATRIX. 2015. **Content Management Directory Matrix** [Online]. Available: <http://www.cmsmatrix.org/> [Accessed 24 diciembre 2015].
- CMS-MATRIX. 2015. Content Management Directory Matrix [Online]. Available: <http://www.cmsmatrix.org/> [Accessed 24 de diciembre de 2015].
- COBREIRO, G. 2015. Informe de Autoevaluación Institucional (2010 - junio/2015). La Habana: Universidad de La Habana.
- CODINA BONILLA, L. & ROVIRA, C. 2006. La web semántica. **Tendencias en documentación digital**. ed. J. Tramullas (Gijón: Trea, 2006): 9-54
- COPPENS, S., MANNENS, E. & VAN DE WALLE, R. 2009. Disseminating heritage records as linked open data. **International Journal of Virtual Reality**, 8, 39.
- CORCHO, Ó., FERNÁNDEZ, M. & GÓMEZ-PÉREZ, A. 2003. Methodologies, Tools and Languages for Building Ontologies: Where is the Meeting Point? **Data & Knowledge Engineering**, 46(1), 41-64.
- CORLOSQUET, S. J. J. 2009. **Bootstrapping the Web of Data with Drupal**. Tesis de maestría, National University of Ireland, Galway.
- CORNELL, U. D. 2009. **VIVO:ISF** [Online]. Available: <https://wiki.duraspace.org/display/VIVO/VIVO-ISF+Ontology> [Accessed 17 de septiembre de 2015].
- CORRENDO, G., SALVADORES, M., MILLARD, I., GLASER, H. & SHADBOLT, N. 2010. Sparql query rewriting for implementing data integration over linked data. In: **EDBT/ICDT Workshop**, New York, NY, USA. ACM, 1 - 11.
- CROFT, W. B. 1987. Approaches to intelligent information retrieval. **Information Processing & Management**, 23, 249-254.
- CUERDA, X. & MINGUILLÓN, J. 2004. Introducción a los Sistemas de Gestión de Contenidos (CMS) de código abierto. **Mosaic**, 36. [Online]. Available: <http://www.uoc.edu/mosaic/articulos/cms1204.html> [Accessed 24 de mayo de 2010].

- CYGANIAK, R. & BIZER, C. 2008. *Pubby-a linked data frontend for sparql endpoints* [Online]. Available: <http://wifo5-03.informatik.uni-mannheim.de/pubby/> [Accessed 17 de enero de 2016].
- DACONTA, M. C., OBRST, L. J. & SMITH, K. T. 2003. *The Semantic Web: a guide to the future of XML, Web services, and knowledge management*, John Wiley & Sons.
- D'ARCUS, B. & GIASSON, F. 2009. *Bibliographic Ontology Specification* [Online]. Structured Dynamics LLC. Available: <http://biblontology.com/> [Accessed 10 de diciembre de 2015].
- DAVIS, M. 2006. Semantic wave 2006: executive guide to the business value of semantic technologies. [Online] USA: Semantic Interoperability Community of Practice, 2006. 56 p. *SICoP white paper series, module, 2*. Available: <http://web-services.gov/SICOPsemwave2006v1.0.doc> [Accessed 5 de septiembre de 2009]
- DE BOER, V., WIELEMAKER, J., VAN GENT, J., HILDEBRAND, M., ISAAC, A., VAN OSSENBRUGGEN, J. & SCHREIBER, G. 2012. Supporting linked data production for cultural heritage institutes: the Amsterdam Museum case study. *The Semantic Web: Research and Applications*. Springer.
- DEAN, M., SCHREIBER, G., BECHHOFER, S., VAN HARMELEN, F., GENDLER, J., HORROCKS, I., MCGUINNESS, D. L., PATEL-SCHNEIDER, P. F. & STEIN, L. A. 2004. OWL web ontology language reference. *W3C recommendation February*, 10. [Online] Available: <http://www.w3.org/TR/owl-ref/> [Accessed 11 de mayo de 2016]
- DÍAZ DOMÍNGUEZ, N. P. 2010. *Los sistemas gestores de contenido y sus aplicaciones*. Tesis de licenciatura, Universidad Veracruzana.
- DIMIĆ SURLA, B., SEGEDINAC, M. & IVANOVIĆ, D. 2012. A BIBO ontology extension for evaluation of scientific research results. In: *Proceedings of the Fifth Balkan Conference in Informatics*. ACM, 275-278.
- DING, Y., MITCHELL, S., CORSON-RIKERT, J., LOWE, B. & HE, B. 2011. The VIVO Ontology: Enabling Networking of Scientists. [Online] Available: <https://pdfs.semanticscholar.org/4047/750dd7b708b0036ff27b0b31c78faf4ec873.pdf> [Accessed 27 de enero de 2016]
- DING, Y., YAN, E., GHAZINEJAD, A. & JIA, H. 2013. Extending the VIVO ontology to iSchools: Enabling networking of information scientists. In: *iConference 2013 Proceedings*.
- DOERR, M., GRADMANN, S., HENNICKE, S., ISAAC, A., MEGHINI, C. & VAN DE SOMPEL, H. 2010. El Modelo de Datos de Europeana (EDM). In: *Meeting: 149. Information Technology, Cataloguing, Classification and Indexing with Knowledge Management*.
- DUCHARME, B. 2013. *Learning Sparql : Queryng and Updating with SPARQL 1.1*, Beijing, O ´Reilly.

- EUROPEANA, F. 2016. *Europeana Collections (version beta)* [Online]. Netherlands. Available: <http://www.europeana.eu/portal/> [Accessed 4 de abril 2016].
- EVANS, D. 2011. Internet de las cosas. *Cómo la próxima evolución de Internet lo cambia todo*. Cisco Internet Business Solutions Group (IBSG) [Online] Available: <http://www.cisco.com/web/LA/soluciones/executive/assets/pdf/internet-of-things-iot-ibsg.pdf> [Accessed 4 de mayo 2015]
- FACULTAD LATINOAMERICANA DE CIENCIAS SOCIALES (FLACSO). 2014. Quiénes somos [Online]. Available: <http://www.flacso.uh.cu/index.php/quienes-somos> [Accessed 13 de enero de 2016].
- FEATHER, J. & STURGES, P. 1997. International encyclopedia of library and information science. London: Routledge.
- FLACSO. 2014. *Facultad Latinoamericana de Ciencias Sociales (FLACSO): Quiénes somos* [Online]. La Habana. Available: <http://www.flacso.uh.cu/index.php/quienes-somos> [Accessed enero 13, 2016 2016].
- FLOREZ FERNANDEZ, H. A. 2013. CONSTRUCCIÓN DE ONTOLOGÍAS OWL. *Vínculos*, 4, 19-34.
- GANDON, F. & SCHREIBER, G. 2014. Rdf 1.1 xml syntax. *W3C recommendation* 25 February 2014. World Wide Web Consortium. [Online] Available: <http://www.w3.org/TR/rdf-syntax-grammar> [Accessed 19 de septiembre de 2015]
- GARCÍA JIMÉNEZ, A. 2004. Instrumentos de representación del conocimiento: tesauros versus ontologías. *Anales de documentación*, 7, 79-95.
- GARCÍA RUIZ, R. 2008. *TFC: XML y Web Semántica*. [Online] Available: <https://core.ac.uk/download/pdf/9627147.pdf?repositoryId=282> [Accessed 4 de mayo 2015]
- GARCÍA, N. E. & CABALLERO, S. D. 2009. Metadatos: necesidad e importancia de integrar estándares. In: *Actas del II Encuentro de Catalogadores*. Biblioteca Nacional de Argentina.
- GARCÍA-GARCÍA, A. 2014. *Integración de contenidos semánticos en un portal web de científicos y humanistas valencianos: Vestigium*. Tesis doctoral, Universidad Católica De Valencia San Vicente Mártir.
- GARCÍA-MARCO, F. J. 2007. Ontologies and knowledge organization: challenges and opportunities for information professionals. *Profesional de la Información*, 16, 541-550.
- GARZÓN, L. 2015. Evolución de la Web. *Tecnología Informática 2015* [Online]. Available: <http://tecnologiainformatica11103.blogspot.com/> [Accessed 19 de marzo 2015].

- GIARETTA, P. & GUARINO, N. 1995. Ontologies and knowledge bases towards a terminological clarification. *Towards very large knowledge bases: knowledge building & knowledge sharing*, 25, 32.
- GILLILAND, A. J. 2008. Setting the Stage. In: BACA, M. (ed.) *Introduction to metadata*. 2nd ed.
- GINGELL, D. 2003. A 15 Minutes Guide to Enterprise Content Management [Online]. Available: http://www.aiim.org.uk/publications/ecm_at_work/pdfs/Ecm_15min_guide.pdf [Accessed 24 de abril de 2013].
- GOLBREICH, C., WALLACE, E. K. & PATEL-SCHNEIDER, P. 2009. OWL 2 Web Ontology Language: new features and rationale. *W3C working draft, W3C (June 2009)* [Online]. Available: <http://www.w3.org/TR/2009/WD-owl2-new-features-20090611> [Accessed 11 de mayo de 2016].
- GÓMEZ-DUEÑAS, L.-F. 2006. La Iniciativa de Archivos Abiertos (OAI): Un nuevo paradigma en la comunicación científica y el intercambio de información. *Revista Códice*, 2, 21-48.
- GÓMEZ-PÉREZ, A., ANGELE, J., LOPEZ, M. F., CHRISTOPHIDES, V., STUTT, A. & SURE, Y. 2002. A survey on ontology tools. [Online]. Available: http://people.aifb.kit.edu/ysu/publications/OntoWeb_Del_1-3.pdf [Accessed 14 de enero de 2009]
- GÓMEZ-PÉREZ, A., FERNÁNDEZ, M. & VICENTE, A. D. 1996. Towards a method to conceptualize domain ontologies. In: "*12th European Conference on Artificial Intelligence (ECAI'96)*", Budapest, Rumanía
- GÓMEZ-PÉREZ, A., FERNÁNDEZ-LÓPEZ, M. & CORCHO, O. 2006. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*, Springer Science & Business Media.
- GRANT, J. & BECKETT, D. 2004. RDF test cases. *W3C recommendation*, 10. [Online]. Available: <http://www.w3.org/TR/rdf-testcases/> [Accessed 19 de septiembre de 2015]
- GRAU, B. C., HORROCKS, I., PARSIA, B., RUTTENBERG, A. & SCHNEIDER, M. 2008. OWL 2 Web Ontology Language: Mapping to RDF Graphs. [Online]. Available: <https://www.w3.org/TR/2008/WD-owl2-mapping-to-rdf-20081202/all.pdf> [Accessed 19 de septiembre de 2015]
- GRAVANO, L., CHANG, C.-C. K., GARCÍA-MOLINA, H. & PAEPCKE, A. 1997. STARTS: Stanford proposal for Internet meta-searching, *ACM SIGMOD Record*, 26(2), 207-218.
- GRUBER, T. 1993a. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *Technical Report KSL- 93-04*. CA: Knowledge Systems Laboratory Stanford University.
- GRUBER, T. 1993b. A translation approach to portable ontologies. *Knowledge Acquisition*, 5, 199-220.

- GRÜNINGER, M. & FOX, M. S. 1995. Methodology for the Design and Evaluation of Ontologies. In: *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing*, IJCAI-95, Montreal.
- GUERRERO BOTE, V. P. & LOZANO-TELLO, A. 1999. Vínculos entre las Ontologías y la Biblioteconomía y Documentación. *In: La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información. Actas del IV Congreso ISKO-España EOCONSID'99*, 22-24 de abril de 1999, Granada. Universidad de Granada, España. 25-31.
- GUZMÁN LUNA, J. A., LÓPEZ BONILLA, M. & DURLEY TORRES, I. 2012. Metodologías y métodos para la construcción de ontologías. *Scientia et Technica*, 2, 133-140.
- GUZMÁN LUNA, J., TORRES PARDO, D., LOPEZ GARCÍA, A. N. 2006. Desarrollo de una ontología en el contexto de la web semántica a partir de un tesoro documental tradicional. *Revista Interamericana de Bibliotecología*, 29 (2), 79-94.
- HASLHOFER, B. & ISAAC, A. 2011. The Europeana Linked Open Data Pilot. *In: International Conference on Dublin Core and Metadata Applications*. The Hague, The Netherlands, 94 - 104.
- HASLHOFER, B. & SCHANDL, B. 2008. The OAI2LOD Server: Exposing OAI-PMH metadata as linked data. *In: International Workshop on Linked Data on the Web (LDOW2008)*, co-located with WWW 2008, Beijing, China.
- HASLHOFER, B. & SCHANDL, B. 2010. Interweaving OAI-PMH data sources with the linked data cloud. *International Journal of Metadata, Semantics and Ontologies*, 5, 17-31.
- HAYES, P. & MCBRIDE, B. 2004. RDF semantics. *W3C recommendation*, 10. [Online]. Available: http://www.open-innovation.in/download/dot/Publication/sub_HaMc2003a.pdf [Accessed 19 de septiembre de 2015]
- HAYES, P. J. & PATEL-SCHNEIDER, P. F. 2014. RDF 1.1 Semantics. *W3C recommendation*, [Online]. Available: <http://www.w3.org/TR/rdf11-mt> [Accessed 19 de septiembre de 2015]
- HEATH, T. & BIZER, C. 2011. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1, 1-136.
- HEFLIN, J. 2007. An Introduction to the OWL Web Ontology Language. Pensilvania: Lehigh University. National Science Foundation (NSF).
- HERNÁNDEZ SAMPIERI, R., FERNÁNDEZ COLLADO, C. & BAPTISTA LUCIO, M. D. P. 2010. *Metodología de la Investigación*, The McGraw-Hill Companies.
- HERNÁNDEZ, G. & PÉREZ, O. 2004. Software propietario y software libre: una cuestión de equilibrio [Online]. Available:

- <http://www.bibliociencias.cu/gsd/collect/eventos/index/assoc/HASH3162.dir/doc.pdf> [Accessed 18 de abril de 2009].
- HERNÁNDEZ-PÉREZ, T. & GARCÍA-MORENO, M. A. 2002. Intranets y preservación digital, algo más que tecnología. *Revista Andaluz de Patrimonio Histórico*, 38, 237 – 243.
- HERNÁNDEZ-PÉREZ, T., RODRÍGUEZ-MATEOS, D., MARTÍN-GALÁN, B. & GARCÍA-MORENO, M. A. 2009. El uso de metadatos en la administración electrónica española: los retos de la interoperabilidad. *Revista española de documentación científica*, 32, 67-91.
- HIDALGO DELGADO, Y. & RODRÍGUEZ PUENTE, R. 2013. La web semántica: una breve revisión. *Revista Cubana de Ciencias Informáticas*, 7, 76-85.
- HIDALGO-DELGADO, Y. 2015. Marco de trabajo basado en los datos enlazados para la interoperabilidad semántica en el protocolo OAI-PMH. Tesis de maestría, Universidad de las Ciencias Informáticas.
- HORRIDGE, M., KNUBLAUCH, H., RECTOR, A., STEVENS, R. & WROE, C. 2004. A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0. *University of Manchester*.
- HORROCKS, I. 2003. Rules for the Semantic Web. *Ieee Intelligent Systems*, 18, 76-77.
- HORROCKS, I., FENSEL, D., BROEKSTRA, J., DECKER, S., ERDMANN, M., GOBLE, C., VAN HARMELEN, F., KLEIN, M., STAAB, S. & STUDER, R. 2000. The Ontology Interchange Language OIL. tech. report, Free Univ. of Amsterdam.
- HORROCKS, I., KRÖTZSCH, M. & PARSIA, B. 2009. OWL 2 Web Ontology Language Conformance. *W3C recommendation*. [Online]. Available: <http://www.w3.org/TR/2009/REC-owl2-conformance-20091027/> [Accessed 27 de agosto de 2013]
- HORROCKS, I., PATEL-SCHNEIDER, P. F. & VAN HARMELEN, F. 2003. From SHIQ and RDF to OWL: The making of a web ontology language. *Web semantics: science, services and agents on the World Wide Web*, 1, 7-26.
- INSTITUTO DE FARMACIA Y ALIMENTOS. 2012. Intranet [Online]. La Habana. Available: <http://intranet.ifal.uh.cu/> [Accessed 15 de enero de 2016].
- IRIS. 2011. *Intranet UH: enlace a Facultades* [Online]. La Habana: Grupo Producciones IRIS. Available: <http://intranet.uh.cu/enlaces/facultades> [Accessed enero 12 2016].
- ISAAC, A. & SUMMERS, E. 2009. SKOS Simple Knowledge Organization System Primer. *W3C Working Group Note*. (2011-11-03). [Online]. Available: <http://dcevents.dublincore.org/index.php/IntConf/dc-2011/paper/view/69/36> [Accessed 4 de abril 2016]

- ISELE, R., JENTZSCH HASSO, A., BIZER, C., VOLZ, J., PETROVSKI, P. & MANNHEIM, U. O. Silk: The Linked Data Integration Framework [Online]. Available: <http://silkframework.org/> [Accessed 12 de diciembre de 2015].
- JCORSON-RIKERT. 2011. VIVO 1.3 Ontology overview diagram [Online]. Universal Wiki Converter. Available: <https://wiki.duraspace.org/display/VTDA/VIVO+1.3+Ontology+overview+diagram> [Accessed 21 de agosto de 2011].
- JIMÉNEZ, M. & MOREO, M. 2006. Diseño de portales mediante CMS. [Online]. Available: <http://lsi.ugr.es/~jparets/Escritos/Dise%F1o%20de%20portales%20mediante%20CMS.pdf> [Accessed 1 de abril de 2009]
- JONES, D. M., BENCH-CAPONAND, T.J.M. & VISSER, P. R. S. 1998. Methodologies for Ontology Development. *IT and KNOWS: Conference of the 15th FIP World Computer Congress*.
- KAMBIL, A. 2008. What is your Web 5.0 strategy?. *Journal of business strategy*, 29, 56-58.
- KARP, P. D., CHAUDHRI, V. K. & THOMERE, J. 1999. XOL: An XML-based ontology exchange language. [Online]. Available: <http://xml.coverpages.org/xol-03.html> [Accessed 11 de febrero 2016]
- KEEFER, A. & GALLARD, N. (eds.) 2007. *La preservación de recursos digitales: El reto para las bibliotecas del siglo XXI*: UOC.
- KERLINGER, L. 2006. *Metodología de la investigación*, McGraw-Hill Interamericana.
- KLYNE, G. & CARROLL, J. J. 2004. Resource description framework (RDF): Concepts and abstract syntax. *W3C recommendation*, 2004. World Wide Web Consortium. [Online]. Available: <http://w3c.org/TR/rdf-concepts> [Accessed 15 de septiembre de 2015].
- KLYNE, G., CARROLL, J. J. & MCBRIDE, B. 2014. RDF 1.1 concepts and abstract syntax. *W3C recommendation*, 25. [Online]. Available: <http://travesia.mcu.es/portalnb/jspui/bitstream/10421/2427/1/RDF%201.pdf> [Accessed 15 de septiembre de 2015].
- KOLLER, A. 2008. Semantic Web and Drupal. *The Semantic Puzzle*. 13 marzo 2008 ed.
- KRSULOVIC-MORALES, E. & GUTIÉRREZ, C. Propuesta para la Creación de una Ontología sobre Departamentos Universitarios de Computación en Chile. Informe de investigación, Universidad de Chile.
- KUNA, H. D., REY, M., PODKOWA, L., MARTINI, E. & SOLONEZEN, L. 2014. Expansión de consultas basada en ontologías para un sistema de recuperación de información. In: *XVI Workshop de Investigadores en Ciencias de la Computación*. (Universidad Nacional de Tierra del Fuego, Argentina). [Online]. Available:

- http://sedici.unlp.edu.ar/bitstream/handle/10915/41513/Documento_completo.pdf?sequence=1 [Accessed 24 de septiembre de 2015].
- LAMARCA, M. J. 2006. *Hipertexto: El nuevo concepto de documento en la cultura de la imagen*. Tesis Doctoral, Universidad Complutense de Madrid.
- LANTHALER, M., & GÜTL, C. (2012). On using JSON-LD to create evolvable RESTful services. In: *Proceedings of the Third International Workshop on RESTful Design* (pp. 25-32). ACM.
- LARIN-FONSECA, R. & GAREA-LLANO, E. 2011. Automatic Generation of the Data-Representation Ontology for Semantic Integration of Heterogeneous Geographical Data. In: *VII Congreso Internacional GEOMATICA*.
- LARRINAGA, F., LIZARRALDE, O., SERNA, A. & GERRIKAGOITIA, J. K. 2013. Caso de uso de open data y linked data en Turismo. *tourGUNE Journal of Tourism and Human Mobility*, 31-39.
- LAZINGER, S. S. 2001. *Digital Preservation and Metadata: History, Theory, Practice*, ERIC.
- LEIVA MEDEROS, A. A. 2011. *Texminer: Un Modelo para el Resumen Automático y la Desambiguación de Textos Científicos en el Dominio de Ingeniería de Puertos y Costas*. Tesis doctoral, Universidad de Granada.
- LEIVA MEDEROS, A. A., SENSO RUIZ, J. A., DOMINGUEZ-VELASCO, S. & HIPOLA, P. 2013. AUTHORIS: a tool for authority control in the semantic web. *Library Hi Tech*, 31, 536-553.
- LESLIE, F. 2015. *Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data*, Apress.
- LIENDL, M., INNOCENTE, V., CASE, M., VAN LINGEN, F., DE ROECK, A., ARCE, P., FÜRTJES, A. & TODOROV, T. 2001. The Role of XML in the CMS Detector Description Database. CERN-CMS-CR-2001-008.
- LLUECA, C. 2003. El proyecto PADICAT (Patrimoni Digital de Catalunya) de la Biblioteca de Catalunya 315 Directrices para la preservación del patrimonio digital [Online] Camberra. Unesco.
- LONGLEY, D., KELLOGG, G., LANTHALER, M. & SPORNY, M. 2015. JSON-LD 1.0 Processing Algorithms and API. *World Wide Web Consortium*. [Online]. Available: <http://www.w3.org/TR/json-ld-api/>. [Accessed 18 de febrero de 2016].
- LÓPEZ, M. F., GÓMEZ-PÉREZ, A., SIERRA, J. P. & SIERRA, A. P. 1999. Building a chemical ontology using methontology and the ontology design environment. *Ieee Intelligent Systems*, 37-46.
- LORENTE, M. 2005. Ontología sobre economía y recuperación de información= Ontology for economics and Information Retrieval.

- Hipertext. Net*, 3. [Online]. Available: <http://www.hipertext.net> [Accessed 23 de abril de 2016].
- LOZANO-TELLO, A., GÓMEZ-PÉREZ, A. & SOSA, E. 2003. Selection of ontologies for the semantic Web. *In*: LOVELLE, J. M. C., RODRIGUEZ, B. M. G., AGUILAR, L. J., GAYO, J. E. L. & RUIZ, M. D. P. (eds.) *Web Engineering, Proceedings*. Berlin: Springer-Verlag Berlin.
- LUKE, S. & HEFLIN, J. 2000. SHOE 1.01. Proposed specification. *Shoe Project*.
- LUNA, J. A. G., PARDO, D. T. & GARCÍA, A. N. L. 2006. *Desarrollo de una ontología en el contexto de la Web semántica a partir de un tesaurus documental tradicional*, Universidad de Antioquia.
- MALMSTEN, M. 2008. Making a library catalogue part of the semantic web. *Metadata for Semantic and Social Applications: Proceedings of the International Conference on Dublin Core and Metadata Applications : Berlin, 22-26 September 2008 : DC 2008* : . Berlin, Germany: Universitätsverlag Göttingen.
- MANOLA, F., MILLER, E. & MCBRIDE, B. 2004. RDF primer. *W3C recommendation*, 10 (6), 1-107. [Online]. Available: <http://www.uazuay.edu.ec/bibliotecas/mbaTI/pdf/RDF%20Primer.pdf> [Accessed 15 de septiembre de 2015]
- MARCOS MORA, M. C. 1999. Los archivos en la era digital. *El Profesional de la Información*, 8(6), 4-13.
- MÁRQUEZ SOLÍS, S. M. 2007. *La Web Semántica*, Chile. [Online] Available: https://books.google.com/cu/books?hl=es&lr=&id=afuncWknStoC&oi=fnd&pg=PA55&dq=SOL%C3%8DS,+S.+M.+2007.+La+Web+Sem%C3%A1ntica,+Lulu.+com.&ots=L6uh1YxKPo&sig=w_wdMGZmXps7bKpHS6EXZAKbYI&redir_esc=y#v=onepage&q&f=false [Accessed 7 de mayo de 2015]
- MARTÍNEZ USERO, J. A. 2004. La necesidad de interoperabilidad de la información en los servicios de administración electrónica: xml, una posible solución. [Online] Available: <http://eprints.sim.ucm.es/5653/1/2004-interoperabilidad.pdf> [Accessed 30 de septiembre de 2015]
- MCGUINNESS, D. L. & VAN HARMELEN, F. 2004. OWL web ontology language overview. *W3C recommendation*, 10(10), 2004. [Online]. Available: http://s3.amazonaws.com/academia.edu.documents/30759881/5.3-B1.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1478816773&Signature=2etuvKxMj2kCFha3ei9d%2FB2M%2B8g%3D&response-content-disposition=inline%3B%20filename%3DOWL_web_ontology_language_overview.pdf [Accessed 17 de febrero de 2016]
- MCKEEVER, S. 2003. Understanding web content management systems: evolution, lifecycle and market. *Industrial Management & Data Systems*, 103, 686-692.

- MÉNDEZ RODRÍGUEZ, E. M. 1999. RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio. *In: Les biblioteques i els centres de documentació al segle XXI: peça clau de la societat de la informació*, 487-498. [Online]. Available: <http://www.cobdc.org/jornades/7JCD/1.pdf> [Accessed 3 de febrero de 2016]
- MICHELINAKIS, D. 2004. *Open Source Content Management Systems: An Argumentative Approach*. Tesis de maestría, University of Warwick.
- MILES, A. & BECHHOFFER, S. 2009. SKOS simple knowledge organization system reference. *W3C recommendation*, 18, [Online]. Available: <http://www.w3.org/TR/skos-reference/> [Accessed 26 de febrero de 2016]
- MILLER, B. & DAVID, D. 2002. *Directions in Web Content Management*. [Online]. Available: <http://www.ddavid.com/groundzero/Directions%20in%20Web%20Content%20Management.pdf> [Accessed 23 de abril de 2014].
- MOTIK, B., GRAU, B. C., HORROCKS, I., WU, Z., FOKOUE, A. & LUTZ, C. 2009a. Owl 2 web ontology language: Profiles. *W3C recommendation*, 27, 61. [Online]. Available: <https://www.w3.org/2007/OWL/draft/ED-owl2-profiles-20090420/all.pdf> [Accessed 17 de febrero de 2016]
- MOTIK, B., PARSIA, B. & PATEL-SCHNEIDER, P. F. 2009b. OWL 2 web ontology language XML serialization. *W3C recommendation, W3C-World Wide Web Consortium*. [Online]. Available: <https://www.w3.org/2007/OWL/draft/ED-owl2-xml-serialization-20090420/all.pdf> [Accessed 17 de febrero de 2016]
- MOTIK, B., PATEL-SCHNEIDER, P. F. & GRAU, B. C. 2009c. Owl 2 web ontology language direct semantics. *W3C recommendation*, 27. [Online]. Available: <https://www.w3.org/2007/OWL/draft/ED-owl2-direct-semantics-20090914/all.pdf> [Accessed 17 de febrero de 2016]
- MOTIK, B., PATEL-SCHNEIDER, P. F., PARSIA, B., BOCK, C., FOKOUE, A., HAASE, P., HOEKSTRA, R., HORROCKS, I., RUTTENBERG, A. & SATTTLER, U. 2009d. OWL 2 web ontology language: Structural specification and functional-style syntax. *W3C recommendation*, 27, 159. [Online]. Available: <http://www.academia.edu/download/42185666/all.pdf> [Accessed 17 de febrero de 2016]
- MUÑOZ, Y. 2015. Evolución de la web y características, Universidad Técnica del Norte. [Online]. Available: <http://es.slideshare.net/AnabelYajaira/evolucion-de-la-web-y-caracteristicas> [Accessed 22 de julio de 2015]
- MYER, T. 2003. *Build an XML-Based Content Management System with PHP* [Online]. Available: <http://www.sitepoint.com/management-system-php/> [Accessed 19 de octubre 2015].

- NAVARRO GALINDO, J. L. & SAMOS, J. FLERSA: Soporte a la Definición de Anotaciones y Búsquedas Semánticas en un CMS. In: *Actas de las XVI Jornadas de Ingeniería del Software y Bases de Datos*, 101. La Coruña, España.
- NAVARRO GALINDO, J. L. 2012. *FLERSA: un sistema semántico de gestión de contenido web (S-CMS)*. Tesis doctoral, Universidad de Granada.
- NECHES, R., FIKES, R. E., FININ, T., GRUBER, T., PATIL, R., SENATOR, T. & SWARTOUT, W. R. 1991. Enabling technology for knowledge sharing. *Ai Magazine*, 12, 36.
- NONAKA, I., TAKEUCHI, H. & KOCKA, M. H. 1999. *La organización creadora de conocimiento: cómo las compañías japonesas crean la dinámica de la innovación*, Oxford University Press México DF.
- NOY, N. F. & MCGUINNESS, D. L. 2001. Ontology development 101: A guide to creating your first ontology. Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA.
- NOY, N. F. & MCGUINNESS, D. L. 2005. Desarrollo de Ontologías-101: guía para crear tu primera ontología. *traducido del inglés por: E. Antezana* [Online] Available: http://protege.stanford.edu/publications/ontology_development/ontology101-es.pdf [Accessed 13 de enero de 2015]
- OJEDA, H. 2014. Mapa Conceptual Evolución de la Web. *Diplomado Componente Docente en Educación Interactiva a Distancia/ Módulo III: Estrategias, Herramientas y Recursos Educativos*. Barquisimeto: Universidad Fermín Toro.
- OSUNA ALARCÓN, M. R. & DE LA CRUZ GÓMEZ, E. 2010. Los sistemas de gestión de contenidos en Información y Documentación. *Revista General de Información y Documentación*, 20, 67-100.
- PAEPCKE, A., BRANDRIFF, R., JANEY, G., LARSON, R., LUDAESCHER, B., MELNIK, S. & RAGHAVAN, S. 2000. Search middleware and the simple digital library interoperability protocol. *DLIB Magazine*, 6.
- PARSIA, B. & SIRIN, E. 2004. Pellet: An owl dl reasoner. In: *Third International Semantic Web* [Conference-Poster] , Hiroshima, Japan.
- PASTOR SÁNCHEZ, J. & DÍAZ ORTUÑO, P. 2008. SPARQL Lenguaje de consulta para RDF. *Retrieved Abril*, 5, 2013. From Recomendación del W3C de 15 de enero de 2008. [Online] Available: <http://skos.um.es/TR/rdf-sparql-query/> [Accessed 3 de abril de 2016]
- PASTOR SÁNCHEZ, J. A. & MARTÍNEZ MÉNDEZ, F. J. 2009. Aplicación de tesauros, taxonomías y ontologías en los sistemas de gestión de contenidos mediante tecnologías de la Web Semántica. *Ibersid*, 3,143-153.
- PASTOR SÁNCHEZ, J. A. 2011a. Drupal 7: Web Semántica al alcance de todos. *Fesabid 2011*.

- PASTOR SÁNCHEZ, J. A. 2011b. *Tecnologías de la web semántica*, Editorial UOC.
- PASTOR SÁNCHEZ, J. A. 2013. Marcado semántico: tecnologías y aplicación para la representación de sistemas de organización del conocimiento en el contexto Linked Open Data. *Scire: representación y organización del conocimiento*, 55-68.
- PASTOR SÁNCHEZ, J. A., ORDUÑA MALEA, E. & SAORÍN, T. 2013. Marcado semántico automático en gestores de contenidos: integración y cuantificación. *El profesional de la información*, 22, 381-391.
- PASTOR-SÁNCHEZ, J. A., MARTÍNEZ-MENDEZ, F. J. & RODRÍGUEZ-MUÑOZ, J. V. 2012. SKOS application for interoperability of controlled vocabularies in the field of linked open data. *Profesional de la Información*, 21, 245-253.
- PEDRAZA JIMÉNEZ, R., CODINA BONILLA, L. & ROVIRA, C. 2007. Web semántica y ontologías en el procesamiento de la información documental. *El Profesional de la Información*, 16, 569-578.
- PEDRAZA-JIMÉNEZ, R., CODINA, L. & ROVIRA, C. 2007. Semantic web and ontologies in document information processing. *Profesional de la Información*, 16, 569-578.
- PÉREZ MATOS, N. E. & SETIÉN QUESADA, E. 2008. Bibliotecología y Ciencia de la Información: enfoque interdisciplinario. *Acimed*, 18(5), 2-26.
- PÉREZ-MONTORO GUTIÉRREZ, M. 2005. Sistemes de gestió de continguts en la gestió del coneixement. *BiD: textos universitaris de biblioteconomia i documentació*, 14 [Online]. Available: <http://bid.ub.edu/14monto1.htm> [Accessed: 10 de noviembre de 2015].
- PINHEIRO, L. V. R. 2006. Ciência da Informação: desdobramentos disciplinares, interdisciplinaridade e transdisciplinaridade. *In*: González de Gómez, M.N.; Dill Orico, E. *Políticas de memória e informação*. Natal: EDUFRRN, 111-142
- POPOV, B., KIRYAKOV, A., OGNYANOFF, D., MANOV, D. & KIRILOV, A. 2004. KIM—a semantic platform for information extraction and retrieval. *Natural Language Engineering*, 10, 375-392.
- PORTELA, P. 2001. Portales, sindicación, contenidos: nuevas oportunidades para los gestores de información. *El Profesional de la Información*, 10, 14-16.
- RAIMOND, Y. & ABDALLAH, S. 2007. The event ontology. [Online] Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.126&rep=rep1&type=pdf> [Accessed 16 de abril de 2016]
- RENSELAER-POLYTECHNIC-INSTITUTE & TETHERLESS-WORLD-CONSTELLATION. 2014. *Vocabularies Encoded in SKOS* [Online]. Available: <https://tw.rpi.edu/web/project/CMSPV/KeyConcepts> [Accessed 28 de noviembre de 2015].

- RÍOS HILARIO, A. B., MARTÍN CAMPO, D. & FERRERAS FERNÁNDEZ, T. 2012. Linked data y linked open data: su implantación en una biblioteca digital. El caso de Europeana. *El Profesional de la Información*, 21 (3), 292-297.
- ROBERTSON, J. 2003. *So, what is a CMS?*. [Online]. Available: http://www.steptwo.com.au/papers/kmc_what/ [Accessed 18 de mayo de 2008]
- RODRIGUEZ GREENBERG, E. M., GEMA BUENO DE LA FUENTE, J., VILA-SUERO, D. & GÓMEZ-PÉREZ, A. 2013. datos. bne. es and MARiMbA: an insight into library linked data. *Library Hi Tech*, 31, 575-601.
- RODRÍGUEZ PEROJO, K. & RONDA LEÓN, R. 2005. Web semántica: un nuevo enfoque para la organización y recuperación de información en el web. *Acimed*, 13(6). Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352005000600003&lng=es&tlng=es [Accessed 7 de mayo de 2015]
- ROSELL LEÓN, Y. 2009. *Impacto de los Sistemas Gestores de Contenido (CMS) en Centros de Educación Superior de Ciudad de La Habana*. Diploma de Estudios Avanzados del Doctorado en Documentación e Información Científica, Universidad de La Habana (Cuba) – Universidad de Granada (España).
- ROSELL LEÓN, Y. 2011. Los CMS y su impacto en Centros de Educación Superior: estudio exploratorio. IV Taller "La Virtualización de la Educación Superior", Evento Provincial UNIVERSIDAD 2012.
- RUIZ JHONES, A. 2016. Entrevista a Directora de Informatización de la Universidad de La Habana[27 de enero de 2016]. La Habana, Cuba.
- SALTON, G. & MCGILL, M. J. 1983. *Introduction to modern information retrieval*, New York.
- SALVADOR BENÍTEZ, A. & RUÍZ RODRÍGUEZ, R. 2005. Metadatos para la preservación de colecciones digitales. *Cuadernos de documentación multimedia*, 16(21), 48-60.
- SÁNCHEZ, J. A. P. 2011. *Tecnologías de la web semántica*, Editorial UOC.
- SANTILLÁN, J. 2007. Introducción a los Sistemas de Gestión de Contenidos [Online]. Available: <http://www.slideshare.net/santillan/introduccion-a-los-sistemas-de-gestion-de-contenidos-cms> [Accessed 15 de enero de 2009].
- SARDUY DOMÍNGUEZ, Y. & URRÁ GONZÁLEZ, P. 2006. Sistemas de gestión de contenidos: en busca de una plataforma ideal. *Acimed*, 14(4) [Online] Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352006000400011&lng=es&tlng=es [Accessed 18 de mayo de 2008].
- SAVOY, J. & PICARD, J. 2001. Retrieval effectiveness on the web. *Information Processing & Management*, 37, 543-569.

- SCHNEIDER, M., CARROLL, J., HERMAN, I. & PATEL-SCHNEIDER, P. F. 2009. OWL 2 Web Ontology Language: RDF-Based Semantics. **W3C recommendation (October 27 2009)**. [Online]. Available: <https://www.w3.org/2007/OWL/draft/ED-owl2-rdf-based-semantics-20090921/all.pdf> [Accessed 17 de febrero de 2016]
- SENSO RUIZ, J. A. 2003. Herramientas para trabajar con rdf. *El Profesional de la Información*, 12, 132-139.
- SERRA SERRA, J. 2002. Estrategias de preservación de documentos electrónicos. In: **V Jornadas de Archivos Electrónicos: El reto electrónico: nuevas necesidades, nuevos profesionales**. Priego de Córdoba: National Archives and Records Administration y el Public Record Office. España.
- SMITH, B., ASHBURNER, M., ROSSE, C., BARD, J., BUG, W., CEUSTERS, W., GOLDBERG, L. J., EILBECK, K., IRELAND, A. & MUNGALL, C. J. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25, 1251-1255.
- SOLOMOU, G. & PAPTAEODOROU, T. 2010. The use of SKOS vocabularies in digital repositories: the DSpace case. In: **Semantic Computing (ICSC)**, 2010 IEEE Fourth International Conference on, IEEE, 542-547.
- SOTO CARRIÓN, J. 2011. **Repositorios Semánticos de Objetos de Aprendizaje**, [Online] Available: https://books.google.com/cu/books?hl=es&lr=&id=-QBxAwAAQBAJ&oi=fnd&pg=PR7&dq=SOTO+CARRI%C3%93N,+J.+2011.+Repositorios+Sem%C3%A1nticos+de+Objetos+de+Aprendizaje,+&ots=0JnzNI902u&sig=708j78cU7k16krPmqS42NOYOFX8&redir_esc=y#v=onepage&q&f=false [Accessed 29 de abril de 2016]
- SOWA, J. F. 2004. Common logic controlled english. Technical report, 2004. Draft, 24 February 2004[Online] Available: <http://www.jfsowa.com/clce/specs.htm> [Accessed 24 de mayo de 2016]
- SPEICHER, S., ARWE, J. & MALHOTRA, A. 2014. Linked data platform 1.0. **Proposed Recommendation, W3C**. [Online]. Available: <http://www.w3.org/TR/2012/WD-ldp-20121025/> [Accessed 5 de Noviembre de 2015].
- SPIVACK, N. 2007. Making sense of the semantic web. [Online]. Available: http://novaspivack.typepad.com/nova_spivacks_weblog/files/nova_spivack_semantic_web_talk.ppt [Accessed 7 de Noviembre de 2015].
- SPORNY, M., LONGLEY, D., KELLOGG, G., LANTHALER, M. & LINDSTRÖM, N. 2014. JSON-LD 1.0. **W3C recommendation (January 16, 2014)**. [Online]. Available: <http://www.w3.org/TR/json-ld/> [Accessed 18 de febrero de 2016].
- STALLMAN, R. 2004. **Software libre para una sociedad libre**, Madrid: Traficantes de Sueños, 2004. Stanford University.

- STEELE, B. 2009. *Cornell's VIVO concept will expand to connect researchers nationwide* [Online]. Available: www.news.cornell.edu/stories/Oct09/VIVOweb.ws.html [Accessed 7 de noviembre de 2015].
- STUDER, R., BENJAMINS, R. & FENSEL, D. 1998. Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering*, 25, 161-197.
- SWARTOUT, B., PATIL, R., KNIGHT, K. & RUSS, T. 1996. Toward distributed use of large-scale ontologies. In: *Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems*, 138-148. Banff.
- TABARES, L. & LEIVA MEDEROS, A. A. 2016. *Consideraciones sobre las estructuras de datos formateadas para las consultas en web semántica*. Tesis de maestría, Universidad de las Ciencias Informáticas, La Habana.
- TALLARICO, M. 2003. *Uso de ontologías en tareas de recupero de información*. Tesina de Licenciatura. [Online] Available: <http://usuarios.fceia.unr.edu.ar/~tallaric/tesina/tesina.pdf> [Accessed 28 de mayo de 2016]
- TAYLOR, R. S. 1966. Professional aspects of information science and technology. *Annual Review of Information Science and Technology*, 1, 15-40.
- TERMENS, M. 2011. Conferencia: Buenas prácticas en preservación digital. (Consideraciones inéditas). Universidad de Barcelona
- TORRE-BASTIDA, A. I., GONZÁLEZ-RODRÍGUEZ, M. & VILLAR-RODRÍGUEZ, E. 2015. Linked open data (LOD) and its implementation in libraries: Initiatives and technologies. *Profesional de la Información*, 24, 113-120.
- TRAMULLAS, J. 2005. Herramientas de software libre para la gestión de contenidos. *Hipertext.net*, 3. [Online] Available: http://www.betsime.disaic.cu/secciones/tec_ja_05.htm [Accessed: 24 de marzo de 2012]
- TRAMULLAS, J. 2006. Intranet: Modelo para el Desarrollo de Redes Corporativas [Online]. Available: <http://www.tramullas.com/pdf/intranet.pdf> [Accessed 16 de mayo de 2013].
- TRAMULLAS, J. 2009. Sistemas de Gestión de Contenido [Online]. Available: <http://www.slideshare.net/tramullas/gestin-de-contenidos> [Accessed 16 de mayo de 2013].
- TRAMULLAS, J. 2011. Los Sistemas de Gestión de Contenidos como pieza fundamental en el despliegue de la Web Semántica. [Online] Available: <http://tramullas.com/los-sistemas-de-gestion-de-contenidos-como-pieza-fundamental-en-el-despliegue-de-la-web-semantica/> [Accessed 24 de febrero de 2014]

- UNIVERSIDAD DE LA HABANA. 2010. *Intranet de la Universidad de La Habana* [Online]. La Habana: Producciones IRIS. Available: <http://intranet.uh.cu/universidad-de-la-habana/mision-vision-y-objetivos> [Accessed 20 de septiembre de 2015].
- URÍAS, G. 2009. *Metodología de la Investigación*, Santa Clara, Universidad Central "Marta Abreu" de las Villas.
- USCHOLD, M. & KING, M. 1995. *Towards a methodology for building ontologies* [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.480.1214&rep=rep1&type=pdf> [Accessed 11 de enero de 2016].
- VAISHNAV, A. & SONWANE, S. S. 2007. Mapping Metadata Standards. *International Conference on Semantic Web and Digital Libraries (ICSD)*. Bangalore.
- VALENCIA CASTILLO, E. 2008. Recuperación y organización de la información a través de RDF usando SPARQL. Universidad Pontificia de Salamanca-Campus de Madrid. [Online]. Available: <http://jena.apache.org/tutorials/sparql.html> [Accessed 15 de septiembre de 2015].
- VAN HEIJST, G., SCHREIBER, A. T. & WIELINGA, B. J. 1997. Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, 46, 183-292.
- VILA-SUERO, D., VILLAZÓN-TERRAZAS, B. & GÓMEZ-PÉREZ, A. 2013. datos. bne. es: A library linked dataset. *Semantic Web*, 4, 307-313.
- VIZCAYA ALONSO, D. (ed.) 1997. *Información: procesamiento de contenido*, Rosario: Nuevo Parahadigma.
- VOLZ, J., BIZER, C., GAEDKE, M. & KOBILAROV, G. 2009. Silk-A Link Discovery Framework for the Web of Data. LDOW, 538. [Online]. Available: <http://vsr-mobile.informatik.tu-chemnitz.de/svnproxy/download/publications/doc/2009/06.pdf> [Accessed 12 de diciembre de 2015].
- W3C 2008. SPARQL Lenguaje de consulta para RDF. **[Entrar como: PASTOR SÁNCHEZ, J. & DÍAZ ORTUÑO, P. 2008...]**
- W3C 2010. Guía Breve de Linked Data.[SI]: El Consorcio. [Online] Available: <http://www.w3c.es/divulgacion/guiasbreves/LinkedData> [Accessed 14 de abril de 2016]
- W3C 2012a. OWL 2 web ontology language document overview. [Online] Available: <https://www.w3.org/TR/owl2-overview/> [Accessed 8 de abril de 2016]
- W3C. 1999. *Resource Description Framework (RDF): Modelo y Sintaxis Especificación* [Online]. Available: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222> [Accessed 28 de septiembre de 2015].

- W3C. 2012b. *SKOS Simple Knowledge Organization System* [Online]. Available: <http://www.w3.org/2004/02/skos/> [Accessed 10 de diciembre 2015].
- W3C. 2013. *W3C Semantic Web Activity* [Online]. Available: <https://www.w3.org/2001/sw/> [Accessed 10 de febrero 2016].
- W3C. 2015. *W3Cwww.w3.org* [Online]. Available: <http://www.w3.org> [Accessed 15 de octubre de 2015].
- WALSH, N. 1998. What is XML. *XML. commune*. [Online] Available: http://www.ce.unipr.it/people/bianchi/Teaching/IntelligenzaArtificiale/rd_f_pl/XML-RDF/xmlguide1.html [Accessed 8 de abril de 2016]
- WILKOFF, N., WALKER, J., ROOT, N. & DALTON, J. 2001. What's next for content management? *Forrester Research Inc., Cambridge*, 4.
- YU, L. 2011. A developer's guide to the semantic Web. *Springer Science & Business Media*. [Online] Available: https://books.google.com/cu/books?hl=es&lr=&id=G0toUPABJEC&oi=fnd&pg=PR7&dq=YU,+L.+2011.+A+developer%E2%80%99s+guide+to+the+semantic+Web,+Springer+Science+%26+Business+Media.+&ots=hyHeGxy2Xz&sig=ccBpEow9dApruMu10qOgh2zUvz8&redir_esc=y#v=onepage&q=YU%2C%20L.%202011.%20A%20developer%E2%80%99s%20guide%20to%20the%20semantic%20Web%2C%20Springer%20Science%20%26%20Business%20Media.&f=false [Accessed 17 de marzo de 2016]
- ZENG, M. & ZUMMER, M. 2009. Introducción de FRSAD y correspondencia con SKOS y otros modelos. In: *World Library and Information Congress, IFLA General Conference and Council*, 75th. Milan. The Hague: IFLA. [Online]. Available: <http://www.ifla.org/files/hq/papers/iflA45/200-zeng-es.pdf> [Accessed 30 de septiembre de 2010].
- ZENG, M. L. & CHAN, L. 2004. Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology*, 55(5), 377-395.

ANEXOS

ANEXOS

Anexo 1: Encuesta 1

CUESTIONARIO

El presente cuestionario se perpetra con el propósito de realizar un estudio para valorar las características semánticas de los CMS (particularmente Drupal y Joomla) y tener conocimiento de los problemas de calidad de los datos generados y gestionados con productos construidos con estos softwares. Necesitamos su más sincera colaboración de forma anónima para lograr un análisis efectivo.

Desde ya agradecemos su colaboración que servirá para mejorar el desempeño y la veracidad de los datos bibliográficos.

Sección 1: Preguntas generales sobre datos del encuestado

1. ¿Qué título o diploma académico obtuvo en el último nivel aprobado?

(Ejemplo: Técnico Medio en Bibliotecología, Lic. en Ciencias de la Información, Ing. Informático, etc.)

2. ¿Posee algún grado científico?

Nota: En caso afirmativo, marque y especifique en qué especialidad (Ejemplo: Especialista, Máster, Doctor, etc.)

_____ Sí
_____ No

2.1 Grado científico:

_____ Especialista
_____ Máster
_____ Doctor

2.2. Especialidad: ***(Ejemplo: Especialista en Recursos Humanos, Máster en Ciencias de la Computación, Doctor en Ciencias, etc.)***

3. Años de experiencia laboral en el trabajo con sitios web construidos con CMS:

_____ Menos de un año
_____ Entre 1 y 5 años
_____ Entre 6 y 10 años
_____ Más de 10 años

4. ¿Qué responsabilidad/cargo u ocupación desempeña en el proceso de creación, implementación y/o desarrollo de los sitios web de la UH?

- _____ Especialista
 _____ Principal
 _____ Especialista
 _____ Técnico
 _____ Otro (Especificar) _____

5. ¿Qué CMS ha utilizado?

- _____ Solo Drupal (Pase a la pregunta 6)
 _____ Solo Joomla! (Pase a la pregunta 7)
 _____ Ambos
 _____ Ninguno (Pase a la Sección 2)

6. Qué versión de Drupal ha trabajado:

7. Qué versión de Joomla ha trabajado:

Sección 2: Conocimiento sobre las características de los CMS

DRUPAL

CMS	Indicadores	Valoración
Drupal	Soporte nativo para la gestión de vocabularios y ontologías	TOMAR EN CUENTA: <i>Capacidad de gestión de RDF y RDFa como formatos nativos para el manejo de vocabularios como bibo, vivo, geontologi, skos, foaf, lom, dublin, core, sioc.</i>
	Asociación entre los contenidos estructurados del portal y los vocabularios RDF gestionados	TOMAR EN CUENTA: <i>Posibilidad de asociar cada tipo estructurado a una o varias clases RDF y mapear parte de los campos del tipo estructurado a propiedades RDF y RDFa.</i>

	Adquisición de módulos	TOMAR EN CUENTA: <i>Posibilidades de extensión e integración a través de la vinculación a módulos extras extendiendo la funcionalidad básica del CMS con posibilidades de acciones semánticas</i>
	Asociaciones RDF	TOMAR EN CUENTA: <i>Posibilidad de publicar las asociaciones RDF creadas en formatos semánticos</i>
	SparQL Endpoints y Linked Data	TOMAR EN CUENTA: <i>Posibilidad de manejar datos abiertos o datos enlazados y de realizar consultas a través de SPARQL Endpoint</i>

JOOMLA

CMS	Indicadores	Valoración
Joomla	Soporte nativo para la gestión de vocabularios y ontologías	TOMAR EN CUENTA: <i>Capacidad de gestión de RDF y RDFa como formatos nativos para el manejo de vocabularios como bibo, vivo, geontologi, skos, foaf, lom, dublin, core, sioc.</i>
	Asociación entre los contenidos estructurados del portal y los vocabularios RDF gestionados	TOMAR EN CUENTA: <i>Posibilidad de asociar cada tipo estructurado a una o varias clases RDF y mapear parte de los campos del tipo estructurado a propiedades RDF y RDFa.</i>
	Adquisición de módulos	TOMAR EN CUENTA: <i>Posibilidades de extensión e integración a través de la vinculación a módulos extras extendiendo la funcionalidad básica del CMS con posibilidades de acciones semánticas</i>

	Asociaciones RDF	TOMAR EN CUENTA: <i>Posibilidad de publicar las asociaciones RDF creadas en formatos semánticos</i>
	SparQL Endpoints y Linked Data	TOMAR EN CUENTA: <i>Posibilidad de manejar datos abiertos o datos enlazados y de realizar consultas a través de SPARQL Endpoint</i>

Anexo 2: Encuesta 2

CUESTIONARIO

El presente cuestionario se perpetra con el propósito de realizar un estudio para valorar las características semánticas de los CMS (particularmente Drupal y Joomla) y tener conocimiento de los problemas de calidad de los datos generados y gestionados con productos construidos con estos softwares. Necesitamos su más sincera colaboración de forma anónima para lograr un análisis efectivo.

Desde ya agradecemos su colaboración que servirá para mejorar el desempeño y la veracidad de los datos bibliográficos.

Sección 1: Preguntas generales sobre datos del encuestado

8. ¿Qué título o diploma académico obtuvo en el último nivel aprobado?

(Ejemplo: Técnico Medio en Bibliotecología, Lic. en Ciencias de la Información, Ing. Informático, etc.)

9. ¿Posee algún grado científico?

Nota: En caso afirmativo, marque y especifique en qué especialidad (Ejemplo: Especialista, Máster, Doctor, etc.)

_____ Sí
_____ No

2.1 Grado científico:

_____ Especialista
_____ Máster
_____ Doctor

9.2. Especialidad: ***(Ejemplo: Especialista en Recursos Humanos, Máster en Ciencias de la Computación, Doctor en Ciencias, etc.)***

10. Conoce al menos 3 sitios web construidos con CMS en la UH:

_____ Si
_____ No

11. ¿Qué responsabilidad/cargo u ocupación desempeña en el proceso de creación, implementación y/o desarrollo de los sitios web de la UH?

- _____ Gestor de contenidos
 _____ Control y seguimiento
 _____ Usuario
 _____ Otro (Especificar) _____

12. ¿Qué CMS conoce?

- _____ Solo Drupal
 _____ Solo Joomla!
 _____ Ambos
 _____ Ninguno

Sección 2: Conocimiento sobre CALIDAD DE LOS DATOS manejados desde Drupal o/y Joomla

¿Cuál es su criterio sobre la capacidad de DRUPAL y /o JOOMLA para el manejo de datos?

Nota: Puede dejar en blanco la casilla en los casos que no conozca. Opcionalmente, puede agregar la opción NEC (NEC: prefiere No Emitir Criterio al respecto)

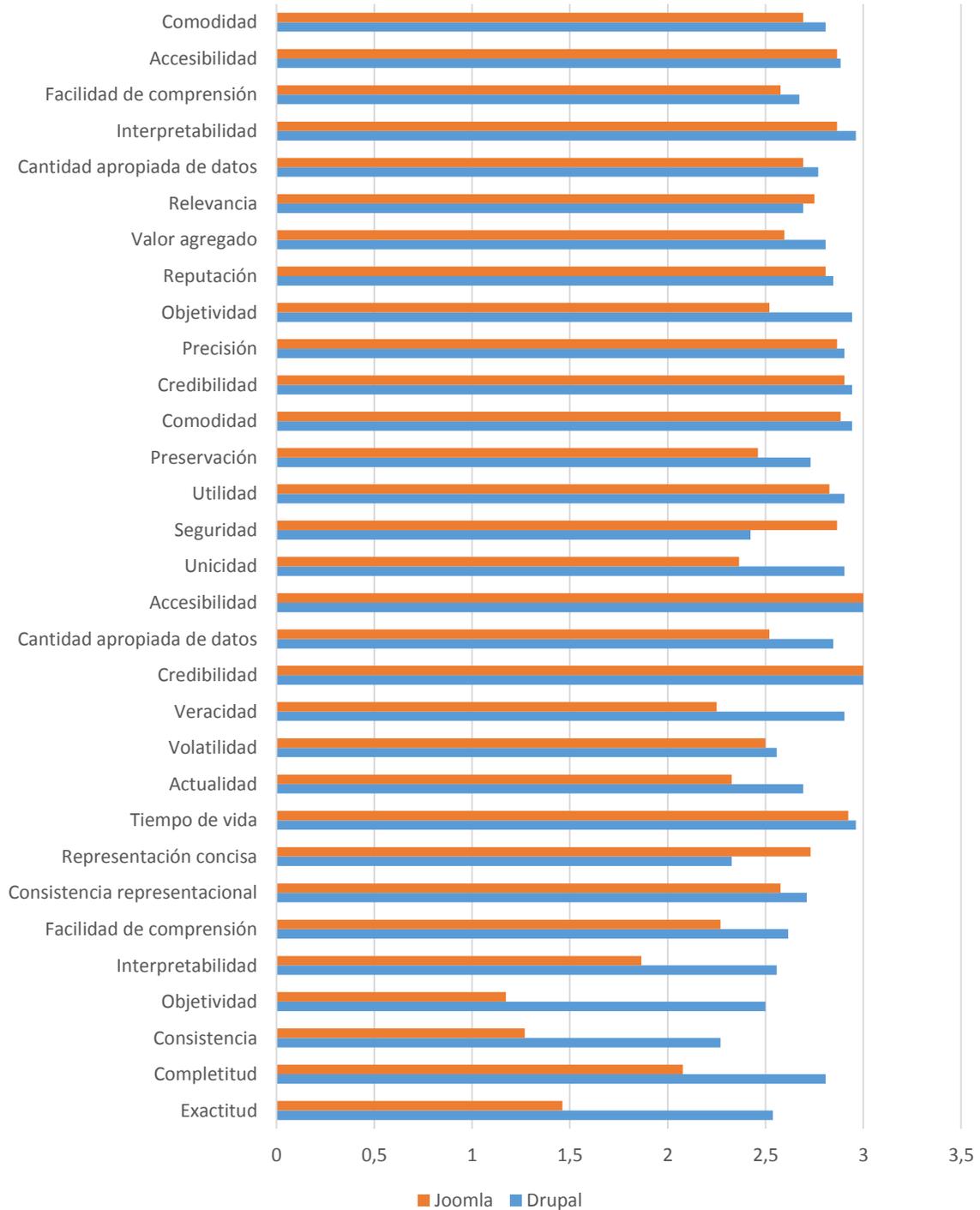
Indicadores	Definición	DRUPAL			JOOMLA		
		baja	media	alta	baja	media	alta
Accesibilidad	Grado en el cual los datos están disponibles o son fáciles y rápidos de recuperar						
Actualidad	Medida de cuan antiguos son los datos, basada en cuánto tiempo hace desde que estos fueron grabados.						
Cantidad apropiada de datos	Grado en el cual el volumen de información es adecuado para una tarea determinada						
Comodidad	Grado en el cual los caminos de navegación resultan difíciles. Número de caminos de navegación perdidos/interrumpidos						
Complejidad	Grado en el cual los datos no se encuentran ausentes o perdidos, cubren las necesidades de las tareas, y son lo suficientemente amplios y profundos para una tarea determinada. No existen valores ausentes para una atributo dado en						

	el sistema						
Consistencia	Grado en el cual la forma y el contenido en los campos de datos es la misma a través de múltiples sistemas fuentes						
Consistencia representacional	Grado en el cual los datos son continuamente presentados en el mismo formato						
Credibilidad	Grado en el cual los datos son considerados como verdaderos y creíbles						
Exactitud	Grado en el cual los datos son certificados, libre de errores, correctos y confiables						
Facilidad de comprensión	Grado en el cual los datos son comprendidos fácilmente						
Interpretabilidad	Grado en el cual los datos están en un lenguaje apropiado y las definiciones son claras						
Objetividad	Grado en el cual los datos son objetivos, sin prejuicios, basados en hechos e imparciales						
Precisión	Número de lugares decimales o dígitos significativos en las medidas. La falta de precisión puede verse como inexactitud.						
Preservación	Propiedad de calidad que poseen los datos digitales. Refleja un estado del dato, el cual puede variar debido a cambios de hardware, software, formatos y procesos a los que el dato es sometido (migración)						
Relevancia	Grado en el cual los datos son aplicables y útiles para una tarea determinada						
Representación concisa	Grado en el cual los datos son representados de forma compacta, bien organizados y estéticamente agradables						
Reputación	Grado en el cual los datos son altamente considerados en términos de sus fuentes o contenidos						
Seguridad	Grado en el cual el acceso a los datos está restringido apropiadamente para						

	mantener su seguridad						
Tiempo de vida	Grado en el cual la edad de los datos es apropiada para una tarea determinada. Tiempo de caducidad del dato. El tiempo de vida tiene dos componentes: la actualidad y la volatilidad.						
Unicidad	Grado en el cual los valores de datos están restringidos en un conjunto con entradas distintas. Cada valor es único en su tipo						
Utilidad	Cada elemento de los datos en el sistema debe satisfacer ciertos requerimientos de las colecciones de los usuarios						
Valor agregado	Grado en el cual los datos son beneficiosos y proporcionan ventajas de su uso. Los datos adicionan valores a las operaciones						
Veracidad	Grado en el cual los datos son inconsistentes, o sea, se refiere directamente a problemas de calidad de datos. Hace referencia al nivel de fiabilidad asociado a ciertos tipos de datos.						
Volatilidad	Medida de cuan inestables pueden ser los datos, o sea, la frecuencia del cambio de un valor para un atributo de entidad						

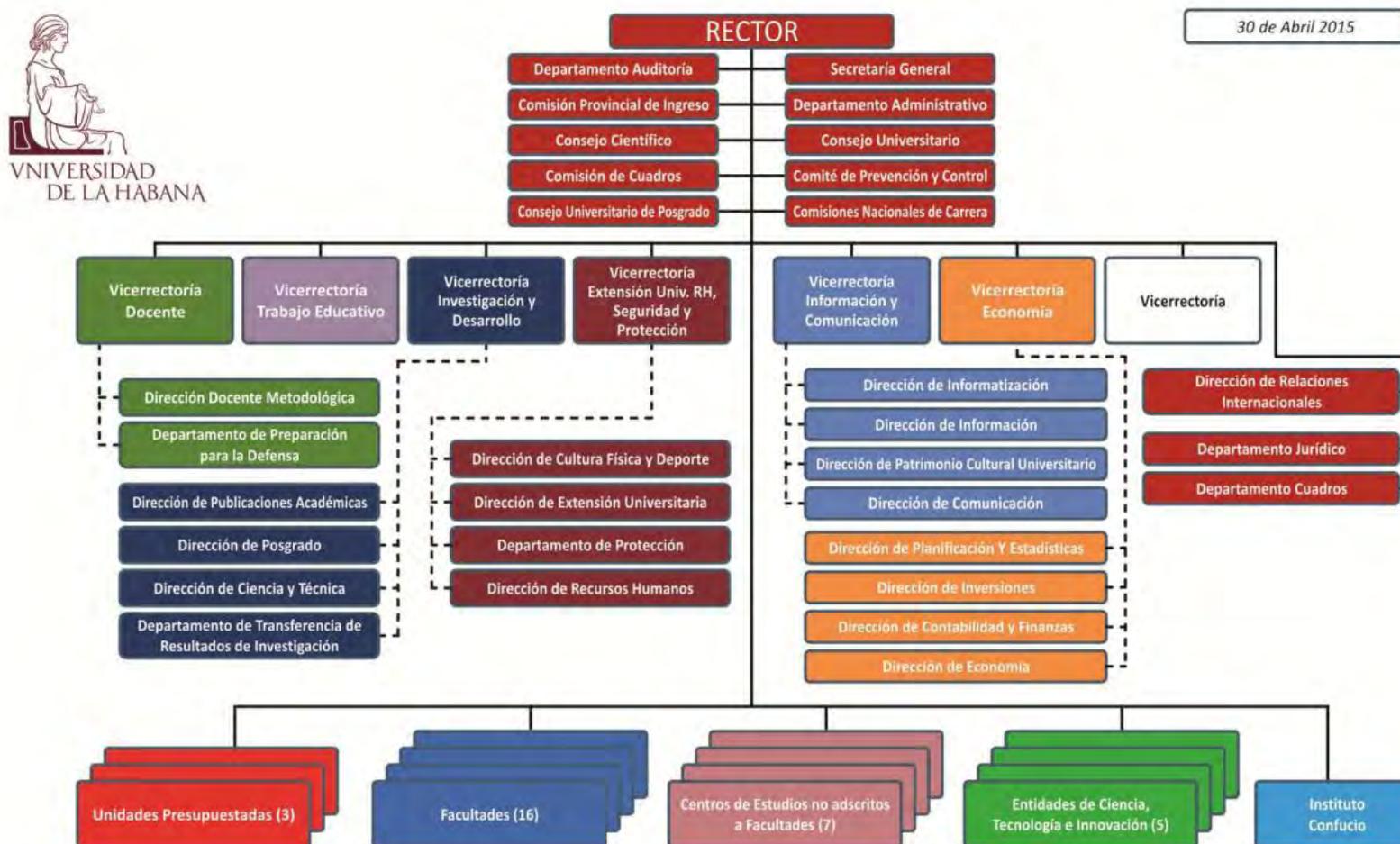
Anexo 3: Comparación entre Joomla y Drupal

Valoraciones sobre manejo de datos en los CMS



Anexo 4: Organigrama de La Universidad de La Habana

Fig. A4.1. Organigrama general



ORGANIGRAMAS DE FACULTADES

Fig. A4.2. Organigrama de la Facultad de Artes y Letras y sus programas académicos

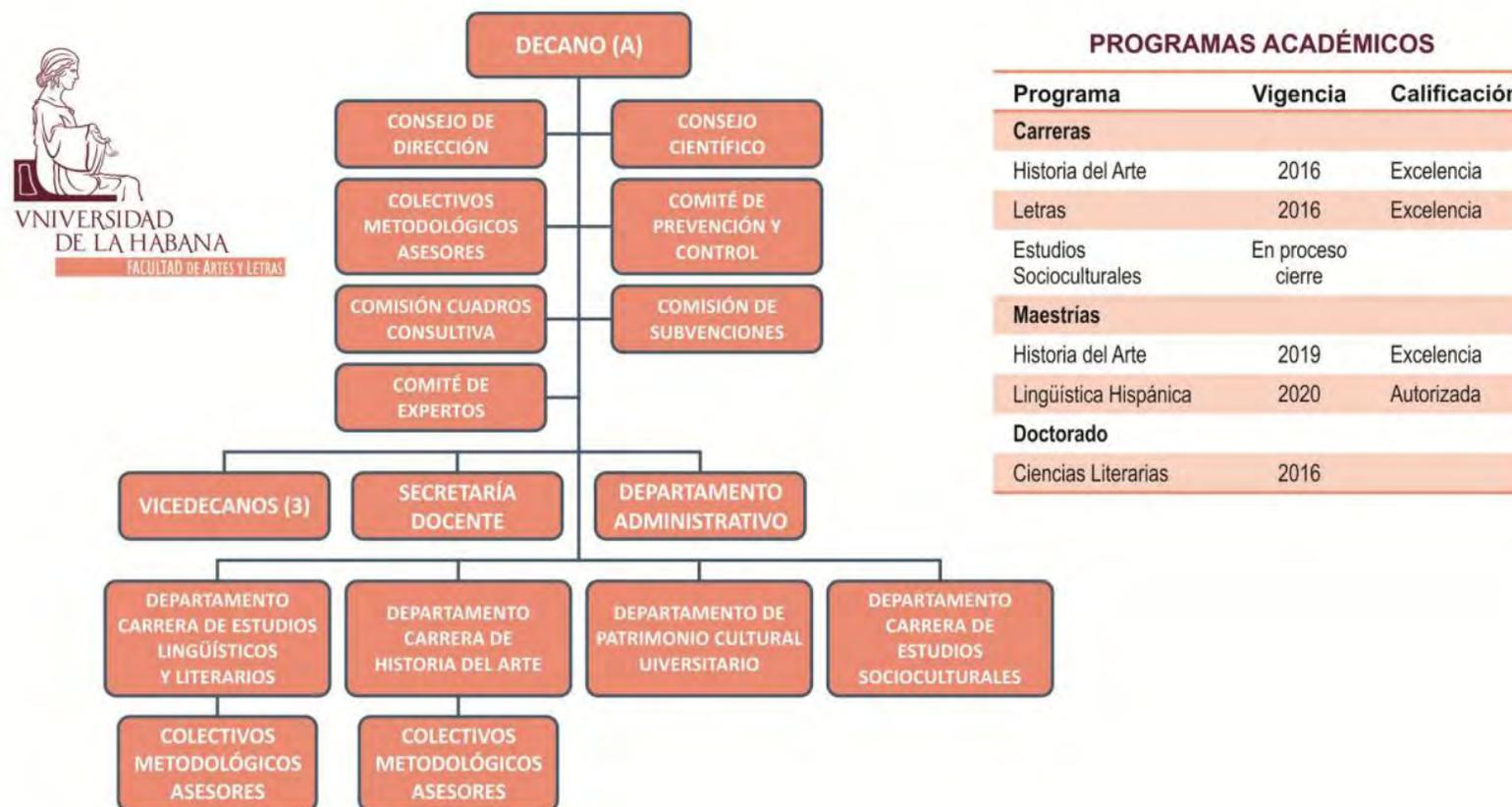


Fig. A4.3. Organigrama de la Facultad de Biología y sus programas académicos

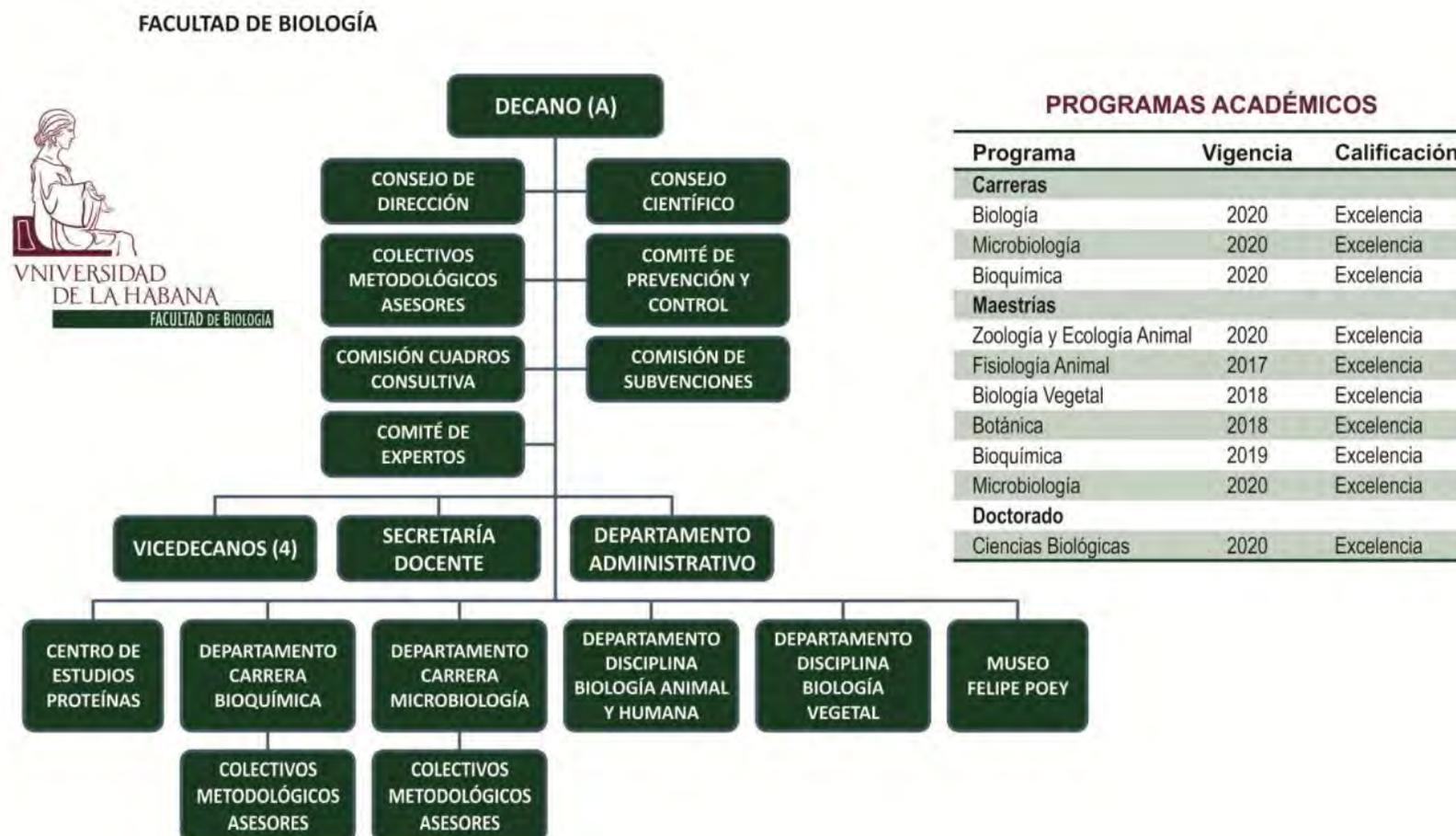


Fig. A4.4. Organigrama de la Facultad de Comunicación y sus programas académicos



Fig. A4.5. Organigrama de la Facultad de Contabilidad y Finanzas y sus programas académicos

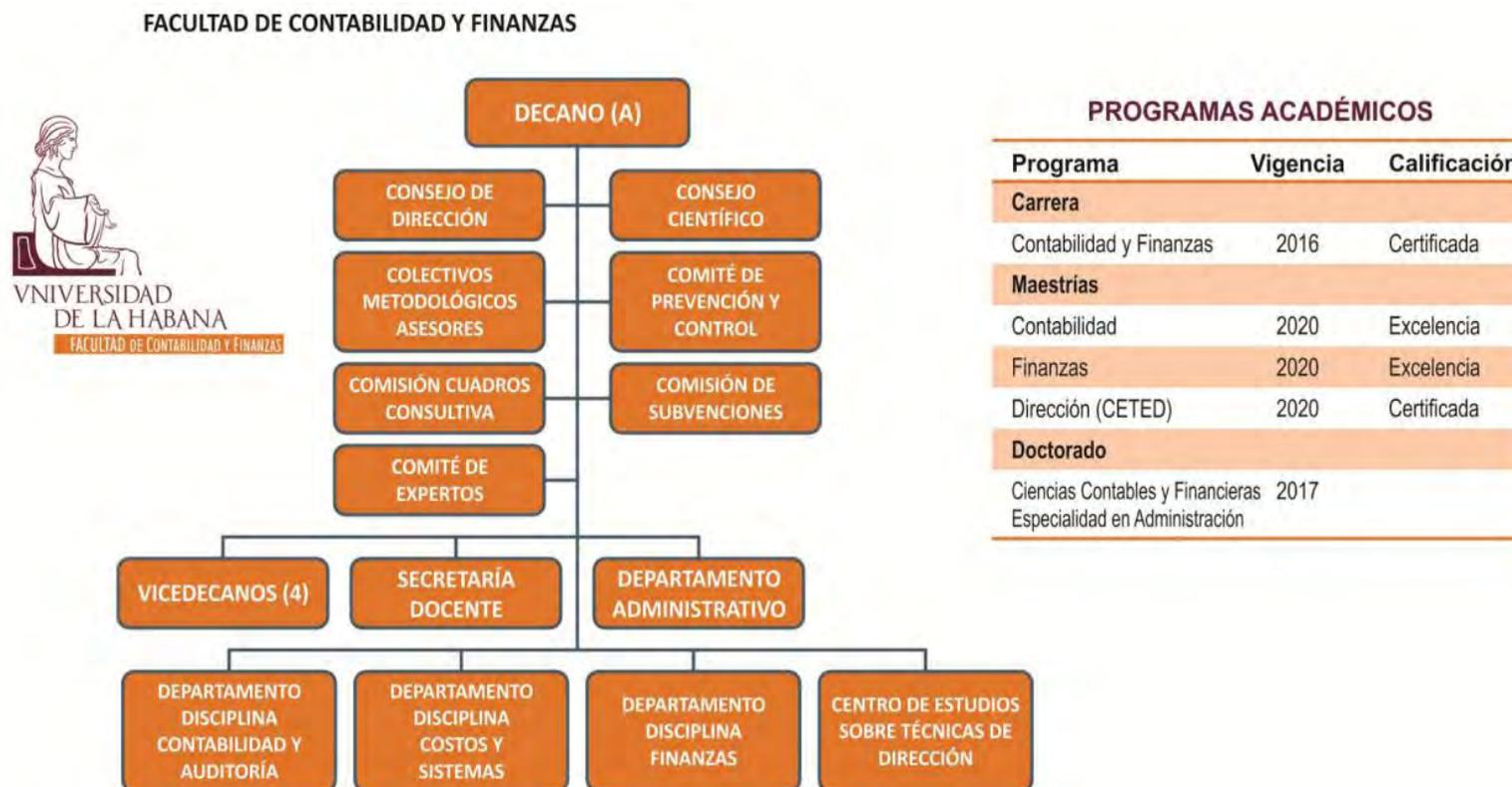


Fig. A4.6. Organigrama de la Facultad de Derecho y sus programas académicos

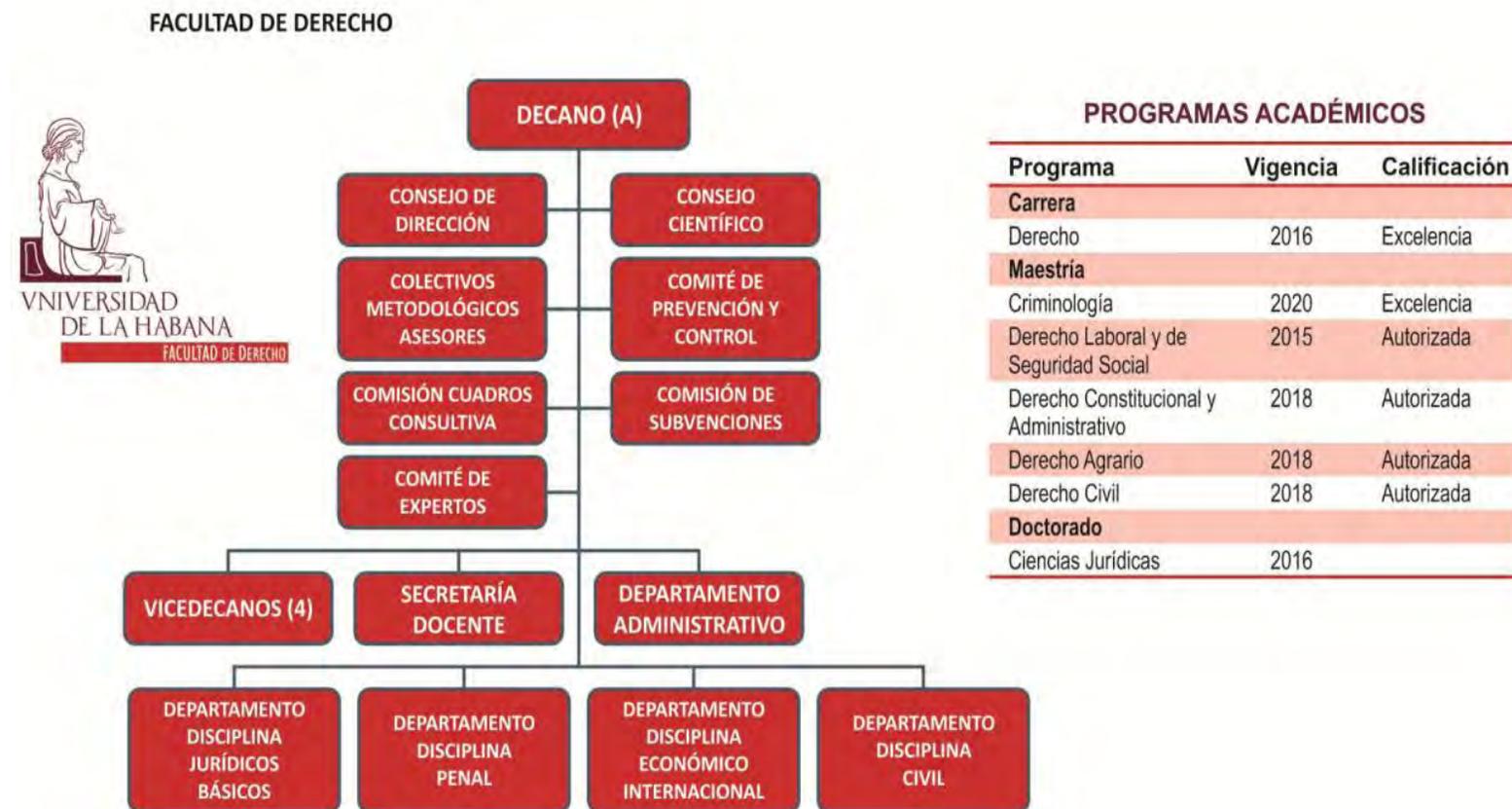


Fig. A4.7. Organigrama de la Facultad de Economía y sus programas académicos

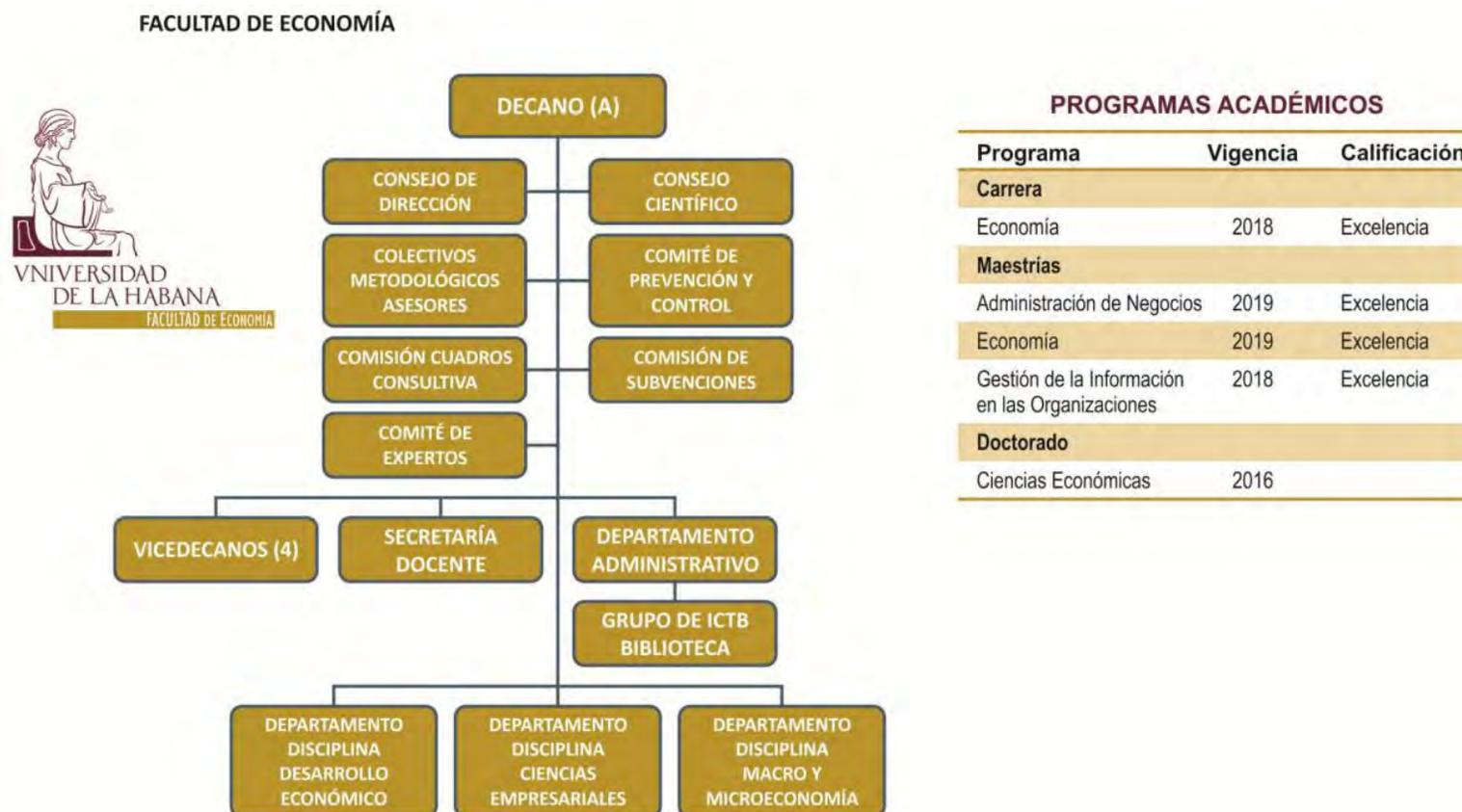


Fig. A4.8. Organigrama de la Facultad de Turismo y sus programas académicos



Fig. A4.9. Organigrama de la Facultad de Filosofía e Historia y sus programas académicos

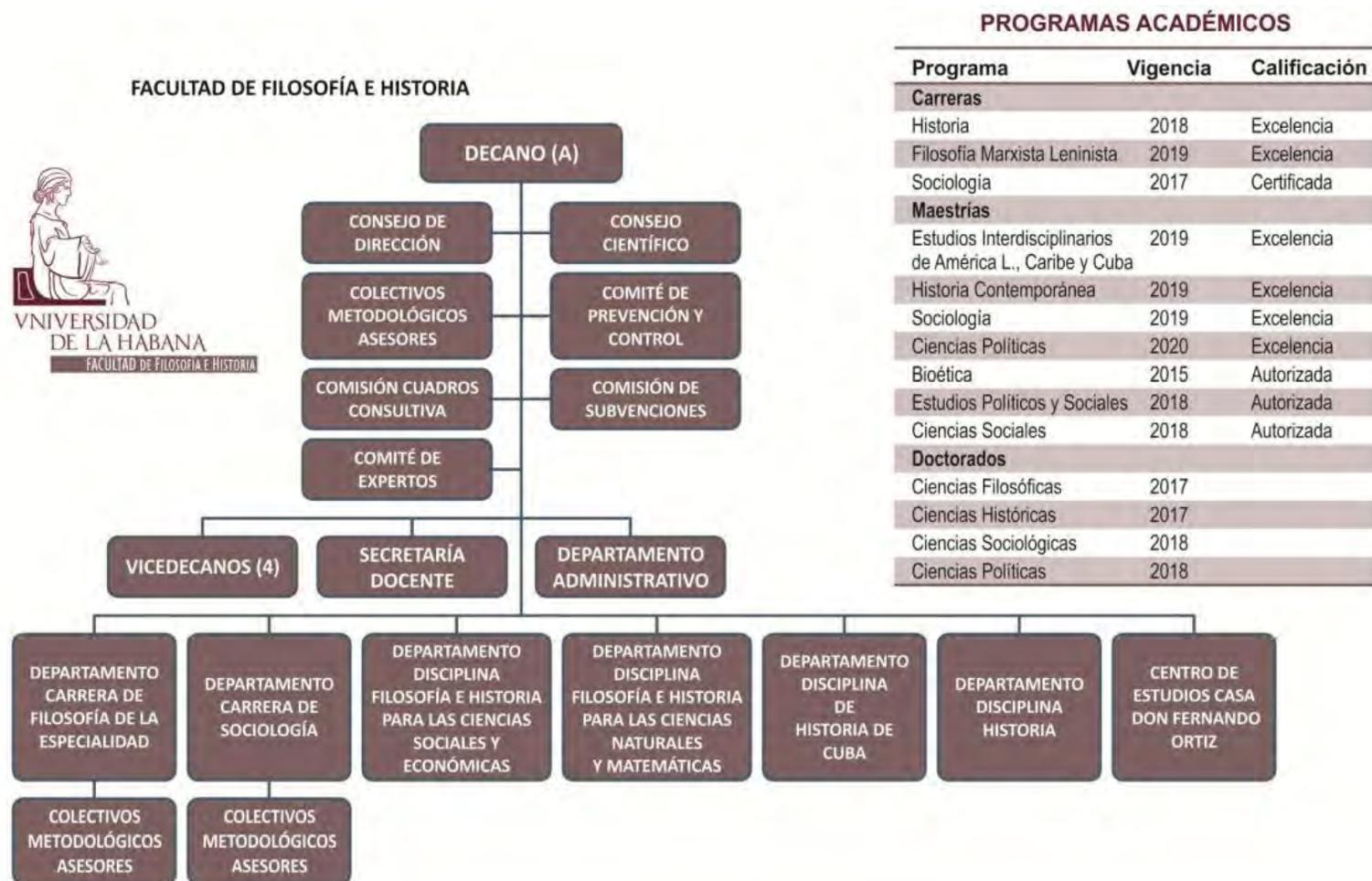


Fig. A4.10. Organigrama de la Facultad de Física y sus programas académicos

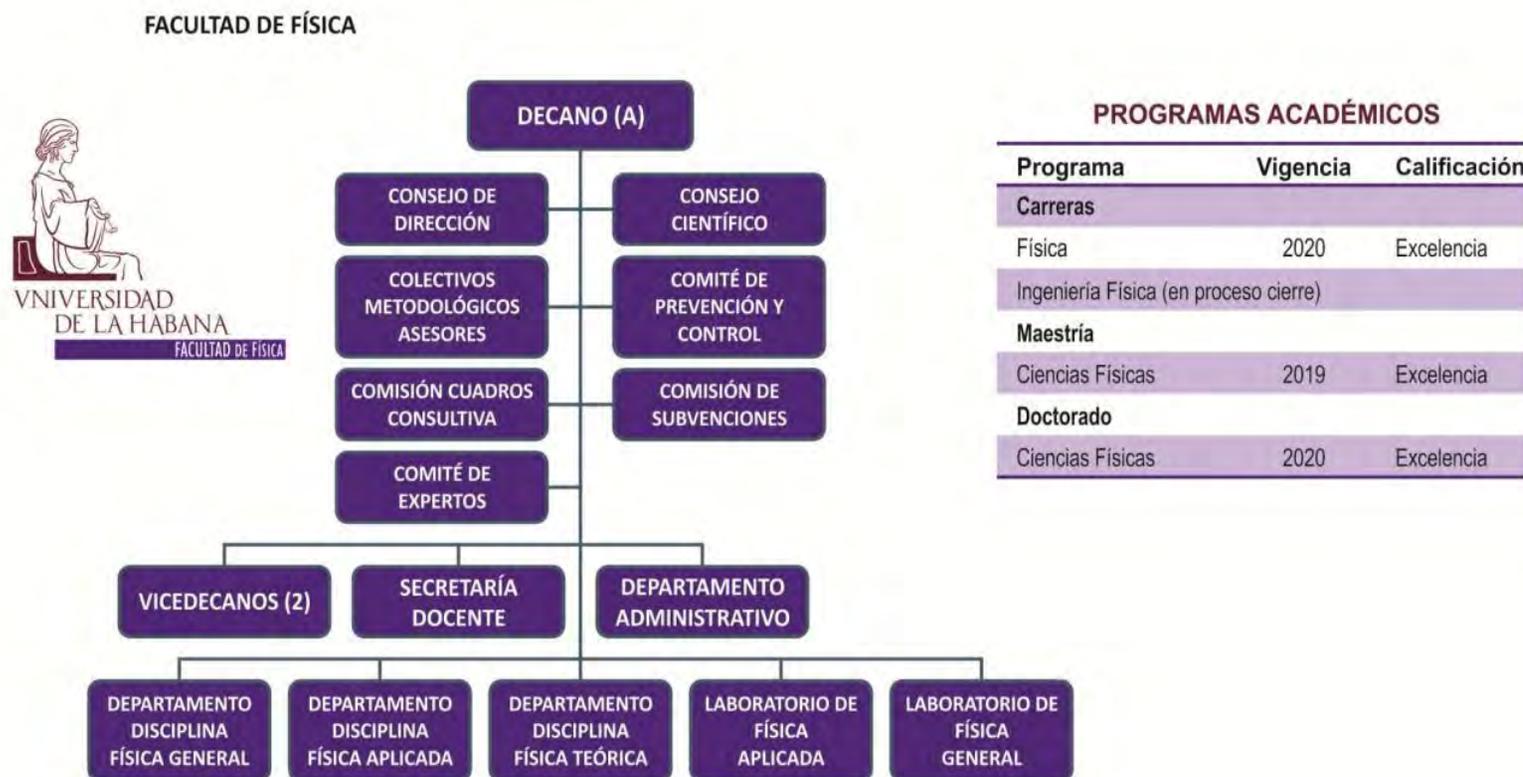


Fig. A4.11. Organigrama de la Facultad de Geografía y sus programas académicos

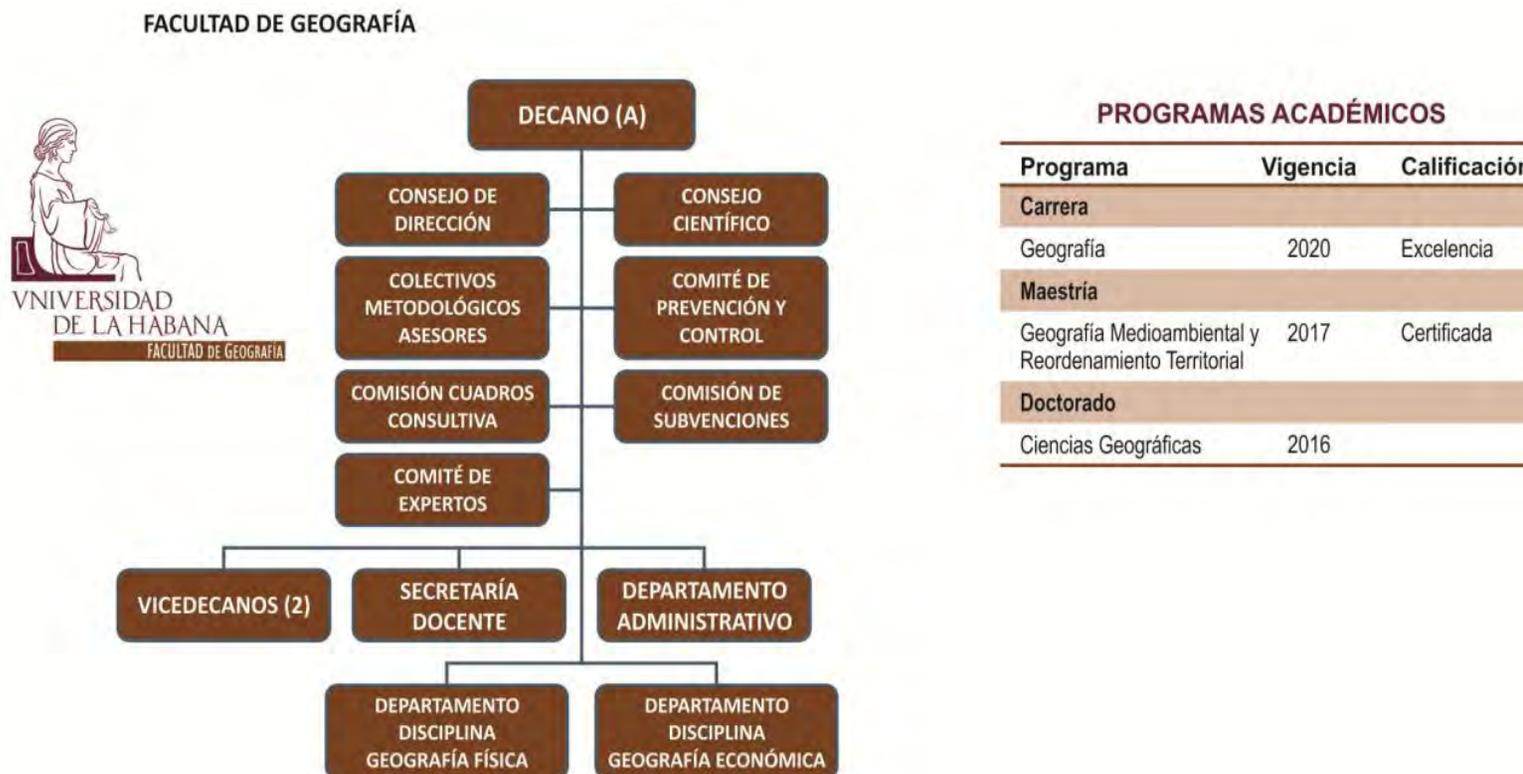


Fig. A4.12. Organigrama de la Facultad de Lenguas Extranjeras y sus programas académicos



Fig. A4.13. Organigrama de la Facultad de Matemática y Computación y sus programas académicos



Fig. A4.14. Organigrama de la Facultad de Psicología y sus programas académicos

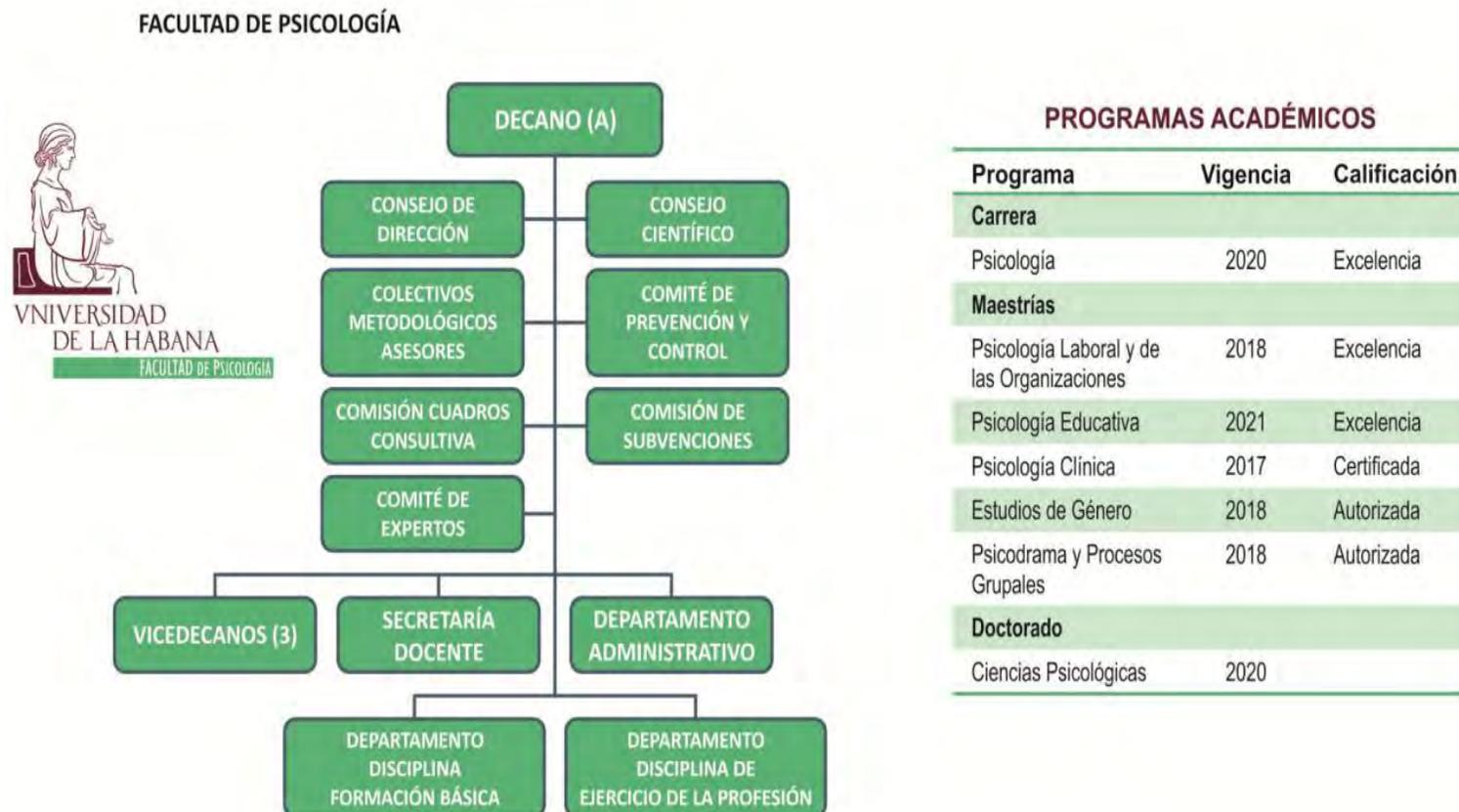


Fig. A4.15. Organigrama de la Facultad de Química y sus programas académicos

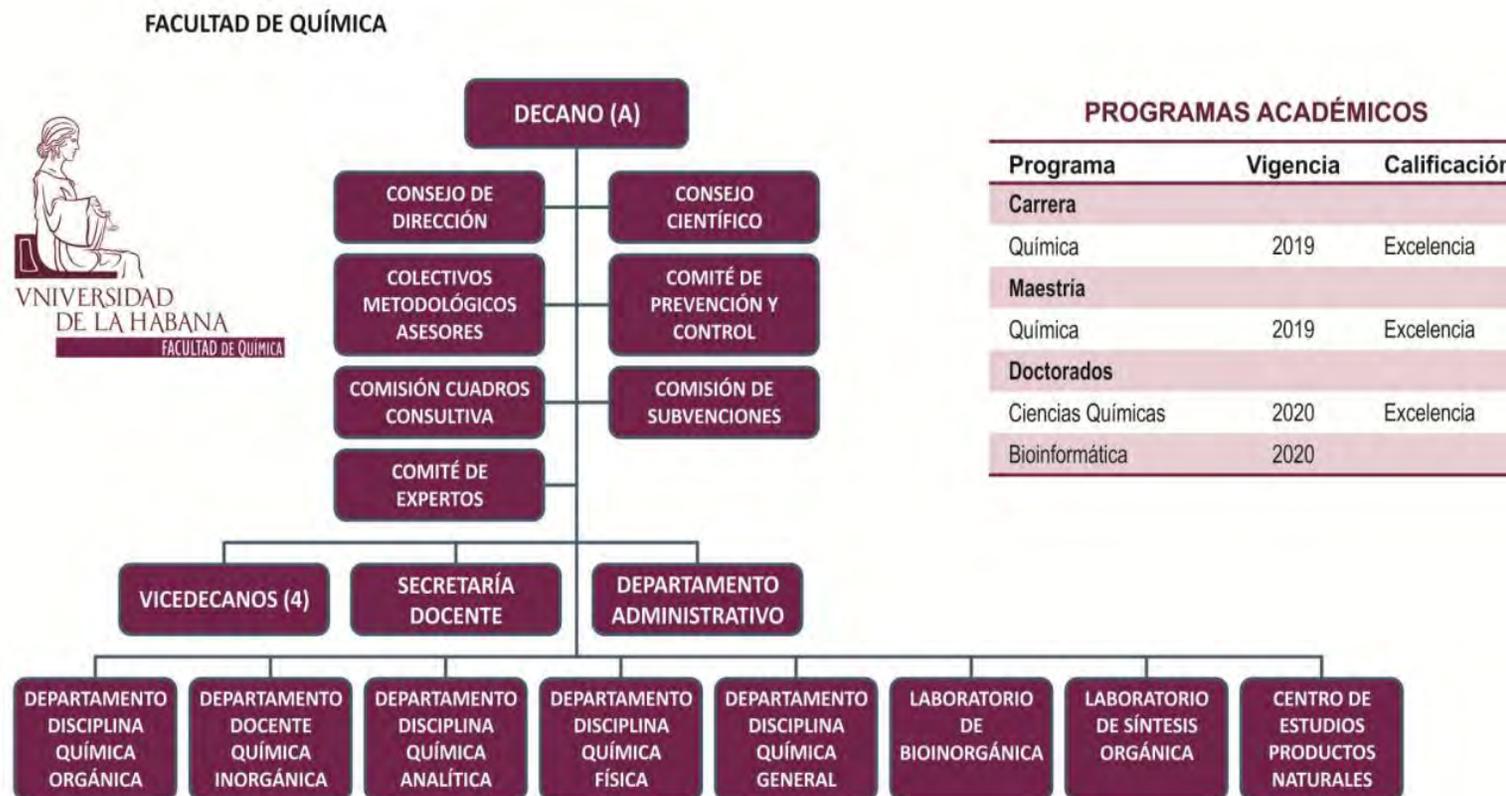


Fig. A4.16. Organigrama de la Facultad de Español para no Hispanohablantes y sus programas académicos

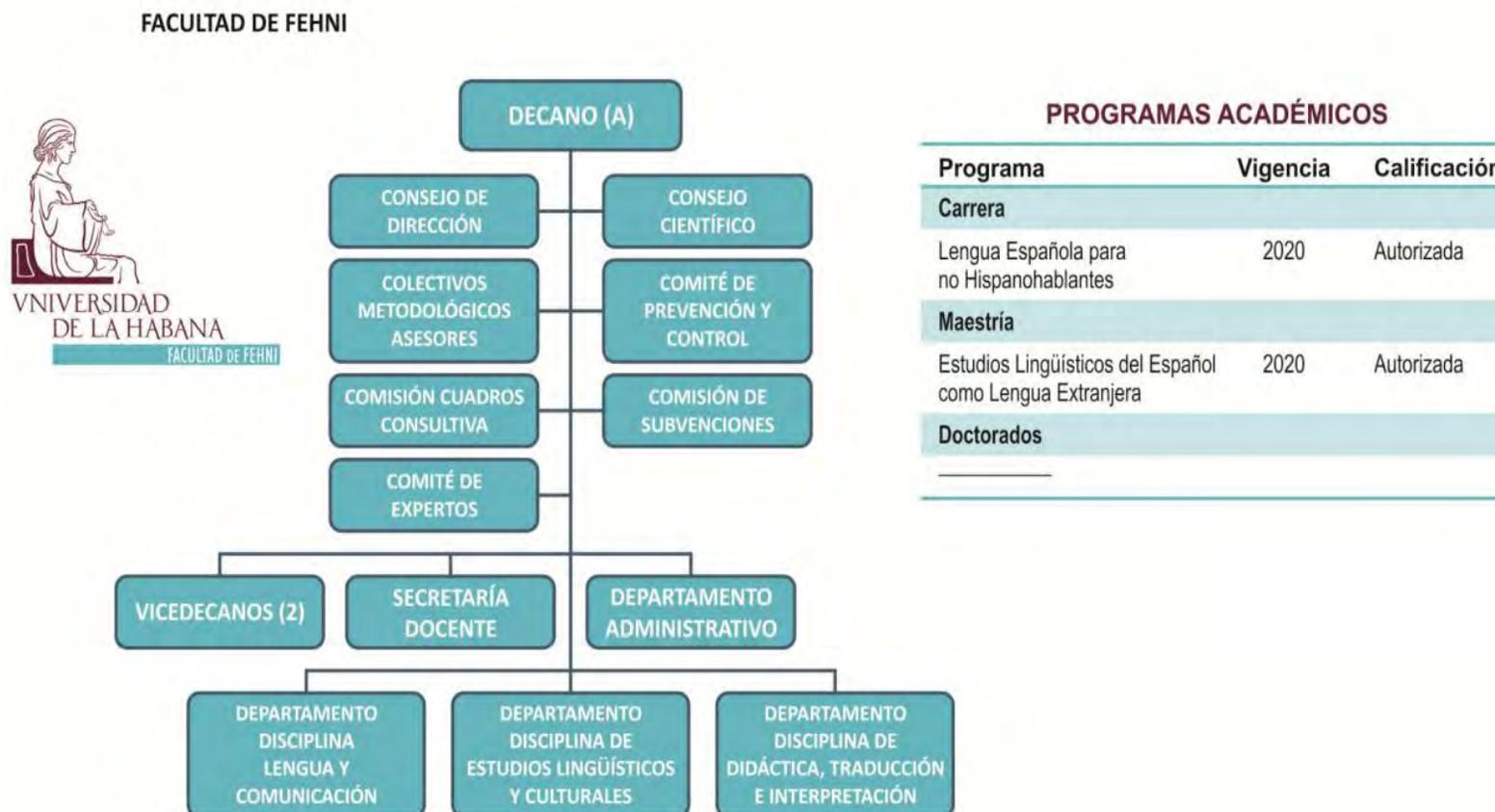
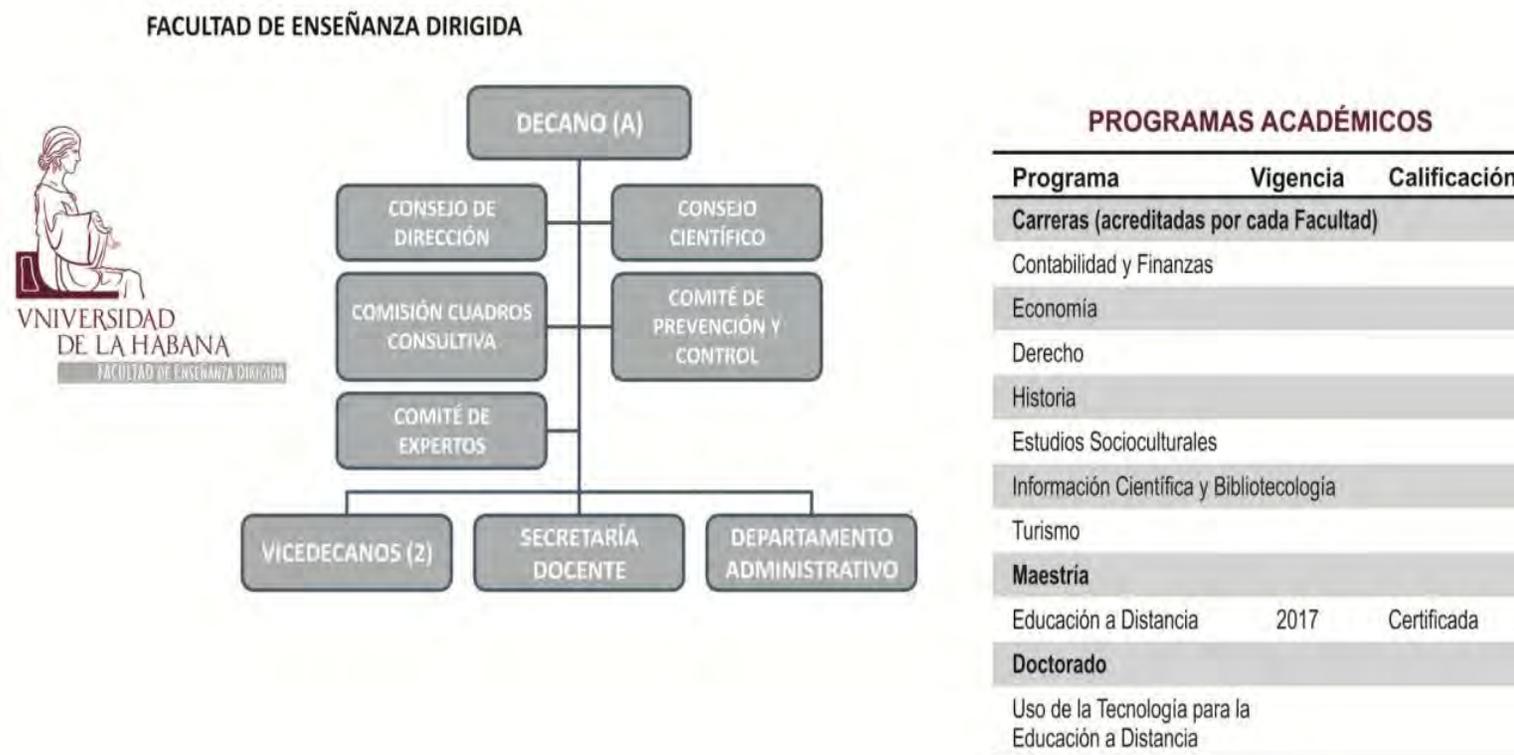


Fig. A4.17. Organigrama de la Facultad de Educación a Distancia y sus programas académicos



ORGANIGRAMAS DE LOS CENTROS DE ESTUDIO NO ADSCRITOS A FACULTADES

Fig. A4.18. Organigrama del Centro de estudios de la Educación Superior (CEPES) y sus programas académicos



PROGRAMAS ACADÉMICOS

Programa	Vigencia	Calificación
Maestría		
Ciencias de la Educación Superior	2019	Excelencia
Doctorados		
Ciencias Pedagógicas	2020	Certificado
Ciencias de la Educación (Es nuevo)		

Fig. A4.19. Organigrama de la Facultad Latinoamericana de Ciencias Sociales(FLACSO)y sus programas académicos



PROGRAMAS ACADÉMICOS		
Programa	Vigencia	Calificación
Maestrías		
Desarrollo Social	2016	Certificada
Gestión y Desarrollo de Cooperativas	2020	Autorizada

Fig. T. Organigrama del Centro de Estudios de Administración Pública (CEAP)y sus programas académicos



PROGRAMAS ACADÉMICOS		
Programa	Vigencia	Calificación
Maestría		
Administración Pública	2016	Autorizada

Fig. A7.20. Organigrama del CEEC**Fig. A7.21. Organigrama del CEHSEU****Fig. A7.22. Organigrama del Instituto Confucio**

ORGANIGRAMAS DE ENTIDADES DE CIENCIA, TECNOLOGÍA E INNOVACIÓN

Fig. A4.23. Organigrama del Instituto de Materiales y Reactivos para la Electrónica (IMRE) y sus programas académicos



PROGRAMAS ACADÉMICOS

Programa	Vigencia	Calificación
Maestría		
Ciencia y Tecnología de los Materiales	2018	Excelencia
Doctorado		
Ciencia de Materiales		

Fig. A4.24. Organigrama de Biomateriales (BIOMAT) y sus programas académicos



PROGRAMAS ACADÉMICOS

Programa	Vigencia	Calificación
Maestría		
Gestión de la Calidad y Ambiental	2019	Certificada

Fig. A4.25. Organigrama del Centro de Estudios de la Economía Internacional (CIEI) y sus programas académicos



Fig. A4.26. Organigrama del Centro de Estudios Demográficos (CEDEM) y sus programas académicos



Fig. A4.27. Organigrama del Centro de Investigaciones Marinas (CIM) y sus programas académicos



ORGANIGRAMAS DE LAS ENTIDADES ADCRITAS A LA UNIVERSIDAD

Fig. A4.28. Organigrama del Jardín Botánico Nacional (JBN) y sus programas académicos

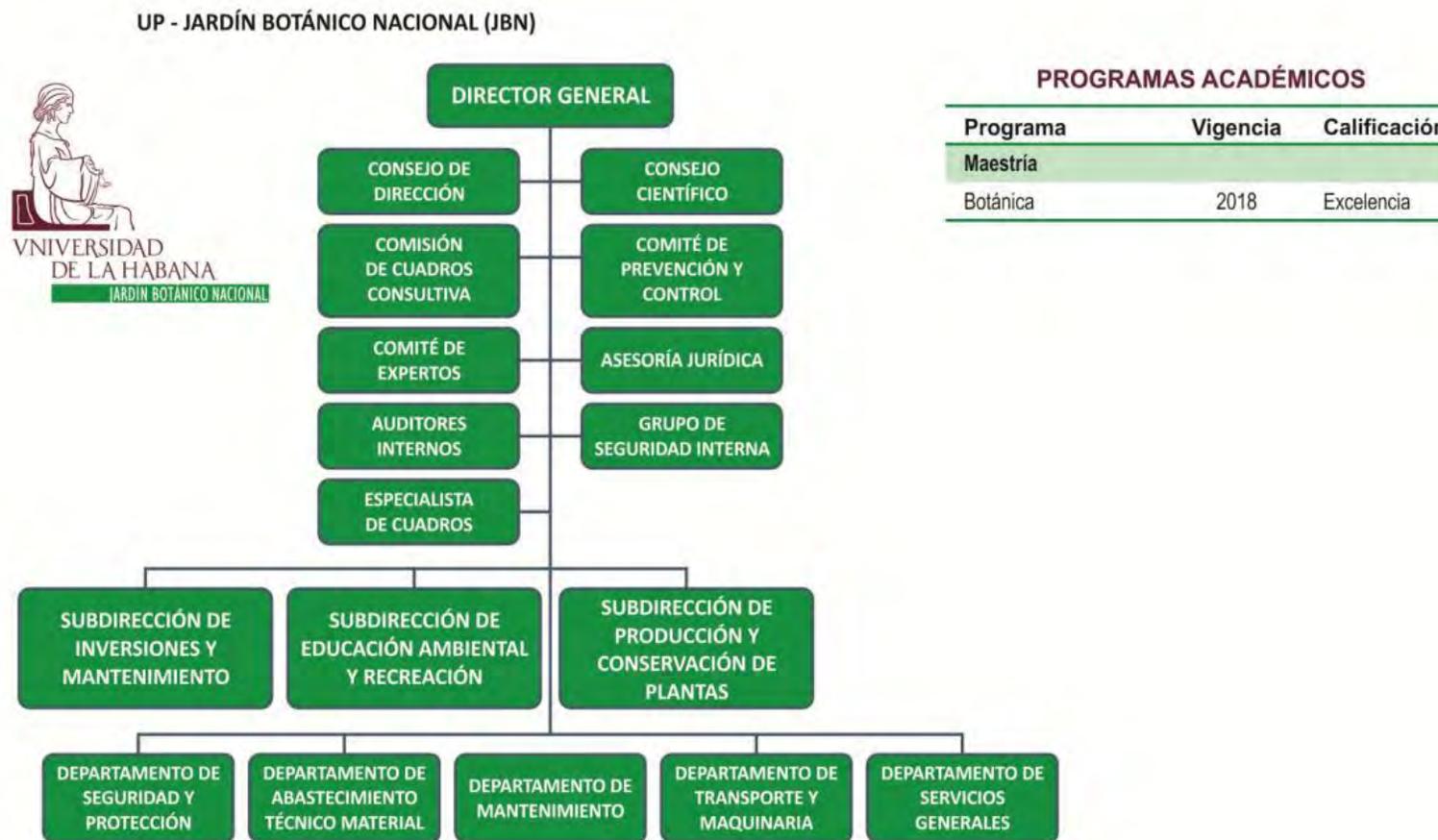


Fig. A4.29. Organigrama del Instituto de Farmacia y Alimentos (IFAL) y sus programas académicos

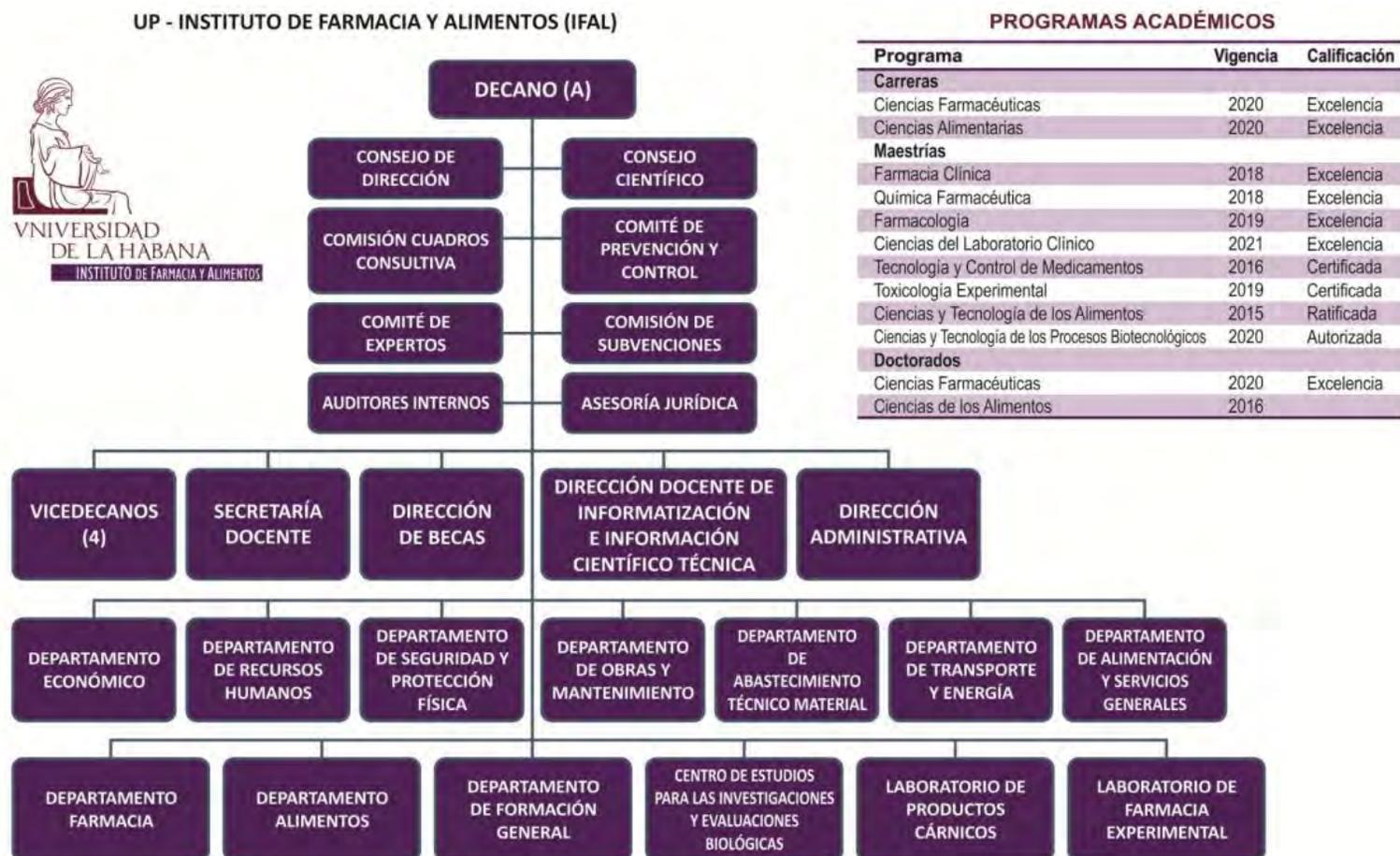


Fig. A4.30. Organigrama de la Unidad Presupuestada de Aseguramiento (UPA)

