

Impacto de los repositorios a través de técnicas cibernéticas: el caso general de Latinoamérica y especial de Costa Rica

Enrique Orduña-Malea

EC3 Research Group. Instituto de Diseño y Fabricación (IDF). Universidad Politécnica de Valencia.

Resumen: A pesar de que las técnicas cibernéticas han sido ampliamente utilizadas en diversas áreas de estudio relacionadas con la actividad científica, tanto en el análisis de entidades (por ejemplo las universidades), como de investigadores, revistas y artículos, su aplicación al análisis de los repositorios es todavía muy escasa, pues el estudio de éstos se ha centrado casi exclusivamente desde la óptica del movimiento *Open Access*, y no desde su carácter de “página web”. En este trabajo se aplican diversos indicadores (topológicos y de mención) al análisis de los repositorios latinoamericanos, así como un análisis específico a nivel de artículo de *Kérwá* (el repositorio institucional de la *Universidad de Costa Rica*). En el caso de los repositorios latinoamericanos, los resultados indican una alta densidad de red en co-menciones entre éstos. En el caso de los artículos depositados en *Kérwá*, los resultados confirman una baja visibilidad web (medida a través de menciones de URL). Se concluye que los indicadores web son útiles para conocer aspectos complementarios del rendimiento de los repositorios, aunque ciertos indicadores (especialmente las menciones de título) deben ser depurados para obtener mejores indicadores a nivel de artículo, así como combinarlos con indicadores de uso.

Palabras clave: Acceso abierto; Repositorios; Cibermetría; Indicadores web; Latinoamérica; Costa Rica; *Kérwá*.

1. Introducción

En la actualidad (octubre de 2013), se cumplen diez años de la celebración de la *Declaración de Berlín*¹ sobre el libre acceso a la literatura científica, y doce años prácticamente de la *Declaración de Budapest* (BOAI)². El repositorio temático *ArXiv*³ comenzó su andadura allá por 1991 (hace más de veinte años) mientras que *RePEc*⁴ lo hizo en 1993. Así mismo, el primer repositorio institucional de la historia (*DSpace*)⁵ comenzó en el *Massachusetts Institute of Technology* (MIT) en 2002. Estos datos demuestran que el *Open Access* (OA) no es un movimiento reciente.

El movimiento Open Access

Desde su aparición y asentamiento (cerca del 50% de los artículos publicados en 2011 se encuentran accesibles)⁶, el movimiento OA ha evolucionado sosteniéndose básicamente en cuatro pilares básicos, que lo definen y caracterizan:

a) Perspectiva social

Las acciones desarrolladas desde esta perspectiva han ayudado a clarificar qué es OA y qué no lo es, así como a explicar las distintas formas de alcanzarlo y a desarrollar estándares y procedimientos para fomentarlo.

De esta forma, las acciones sociales se podrían clasificar, básicamente, en la celebración de diversas Declaraciones (como las de Budapest, Bethesda, Berlín, o más recientemente de Alhambra⁷), así como en el diseño y establecimiento de diversas estructuras, entendiendo por ello la materialización política e institucional

del paradigma OA obtenido de las distintas Declaraciones y encuentros de la comunidad.

b) Perspectiva legal

El enfoque legal se ha dirigido a estudiar los aspectos relacionados con los derechos de autor de las publicaciones, los conceptos de *copyright* y *copyleft*, así como el desarrollo y aplicación de diversas licencias (como las *Creative Commons*⁸ o *Science Commons*⁹). Igualmente, los aspectos relativos a los mandatos de apertura en las instituciones o las iniciativas para depositar en repositorios los resultados de las investigaciones financiadas con fondos públicos (donde resultan paradigmáticos casos como el *National Institutes of Health* en Estados Unidos, entre muchos otros) han sido objetos de interés por parte de este importante pilar del movimiento OA.

c) Perspectiva técnica

Igualmente, el desarrollo informático y técnico ha sido fundamental para que tanto la ruta dorada (revistas) como la ruta verde (repositorios) hacia el OA fueran posibles. Los desarrollos se centran fundamentalmente en:

- El software necesario para gestionar la documentación en abierto, tanto para las revistas electrónicas (como *OJS*) como para los repositorios (*DSpace*, *EPrints*, *Fedora*, etc.).
- El desarrollo de estándares y protocolos de intercambio. Tanto los pioneros (OAI-PMH y OAI-ORE), como otros que fueron desarrollándose posteriormente (SWORD, LOCKSS, etc.).

d) Perspectiva documental

Finalmente, el enfoque documental se centra en la gestión de los documentos depositados en las distintas plataformas (tanto repositorios como revistas), y se enfocan principalmente en la descripción de los recursos (metadatos) y en su almacenamiento y recuperación, tanto interna (dentro de la plataforma) como externa (desde los motores de búsqueda u otros directorios o servicios).

Por tanto, es la unión de estas cuatro perspectivas la que permite al movimiento OA crecer y evolucionar, tanto a nivel profesional (nuevos proyectos, servicios, etc.) como académico (posibles ventajas del OA en la citación o los cambios en el comportamiento de los científicos, etc.).

Importancia de los repositorios para el OA

Los “vehículos” creados para llevar a la práctica el movimiento OA, tal y como se ha comentado anteriormente, son las revistas y portales de acceso abierto (la llamada ruta dorada) y los repositorios (ruta verde), siendo éstos últimos el objeto de este trabajo.

Las principales funciones de los repositorios, ya ampliamente conocidas, son las de

- servir de mecanismo de depósito: facilitar el autoarchivo para preservar el acervo científico, y
- servir de mecanismo de acceso: facilitar los procedimientos de recuperación de información.

Los repositorios pueden clasificarse en las siguientes tipologías:

- a) Documentales: centrados en estructuras intelectuales determinadas (tesis doctorales, artículos, posters, etc., y más recientemente datos brutos de investigación).
- b) Temáticos: basados en los contenidos producidos por un área específica del conocimiento.
- c) Institucionales: centrados en los documentos producidos por una (o varias) instituciones u organizaciones.

Dejando aparte los repositorios basados en tipologías documentales concretas (muy minoritarios), los repositorios temáticos e institucionales surgen y se desarrollan con objetivos y funciones claramente diferentes.

Por un lado, los repositorios temáticos surgen debido a la existencia de una cultura de comunicación informal entre científicos (un claro ejemplo de ello lo constituye *ArXiv*), mientras que la principal función de los repositorios institucionales es la de registrar la actividad académica de una o varias instituciones u organizaciones.

Los repositorios institucionales juegan por tanto un papel fundamental en la creación del rastro e identidad digital de la universidad en la Red (Aguillo, 2009). Precisamente, la consideración del repositorio como depósito y servicio de difusión y acceso a su producción académica (y por tanto un reflejo de sus actividades), lleva a la mayor parte de las mejores universidades del mundo a crear sus propios repositorios institucionales durante la primera década del siglo XXI.

No obstante, supone un error creer que el servicio del repositorio institucional acaba cuando éste es creado y alimentado con una carga inicial de ítems, ya que “la finalidad de todo repositorio institucional es la difusión y aumento del impacto entre la comunidad científica de los resultados de investigación, evitando convertirse en un nuevo depósito infrutilizado y carente de sentido” (Sánchez y Melero, 2006). La mayoría de iniciativas y nuevas Declaraciones carecen actualmente de sentido si se limitan a ser meras expresiones de apoyo a una iniciativa ya consolidada y no revierten en el desarrollo de nuevos servicios sobre la capa de infraestructura ya creada (los repositorios), que permitan atender las necesidades de los investigadores de una manera más adecuada (De-Castro, 2014).

Entre los principales problemas de infrutilización de los repositorios, se pueden distinguir los siguientes:

- Colecciones incompletas o desactualizadas.
- Carencia de servicios avanzados en la plataforma.
- Navegabilidad y arquitectura de información inadecuados.
- Indización insuficiente o limitada en los motores de búsqueda.

Mientras que el primer punto se debe principalmente al abandono de las plataformas una vez son lanzadas oficialmente, debido al escaso interés en mantenerlas actualizadas o a la carencia de recursos económicos y humanos, los restantes puntos vienen determinados por una visión demasiado estrecha de lo que la existencia de un repositorio institucional supone.

Mientras todo el énfasis en la creación, diseño y difusión del producto toma como centro de gravedad el concepto “Open Access”, generalmente se olvida que el producto es, ni más ni menos, que una sede web, y por tanto se deben cuidar ciertos aspectos relacionados con la navegabilidad, la usabilidad, la visibilidad en los motores de búsqueda, etc.

Todos estos asuntos quedan normalmente de lado al confiarlos casi completamente al software utilizado para gestionar la colección, y cuya configuración por defecto no suele ser la más adecuada para una página web.

Una de las principales consecuencias es una mala experiencia de usuario al navegar e interactuar en el repositorio, así como las bajas tasas de indización que los contenidos de estos repositorios tienen en los motores de búsqueda (especialmente *Google Scholar*).

En ese sentido, merece especial atención el trabajo de Arlitsch et al., 2012), quienes identifican una ratio de indización de los repositorios institucionales en *Google Scholar* muy bajo. Las principales razones identificadas son las pocas facilidades de navegación para los robots de búsqueda y el esquema de metadatos utilizado (generalmente *Dublin Core*), no adecuado para los requisitos de *Google Scholar*.

Las bajas ratios de indización en *Google* son de suma importancia si se tiene en cuenta que éste es actualmente uno de los puntos clave de comienzo de búsqueda de información científica por parte de los usuarios. Ya en el año 2005, la OCLC alertaba de que el 89% de los estudiantes comenzaban sus búsquedas en los motores de búsqueda, y sólo un 2% lo hacía a partir de las sedes web de las bibliotecas académicas (DeRosa y OCLC, 2005). Informes más recientes siguen alertando que los usuarios encuentran la información de forma mucho más sencilla a través de *Google* que de las propias páginas universitarias¹⁰, lo que constituye un auténtico hándicap para la visibilidad y uso de los repositorios institucionales.

Dado el alto uso de los motores de búsqueda como punto de partida de las consultas, especialmente las académicas (y que en la mayoría de las ocasiones se accede directamente al recurso sin pasar por la página web del repositorio), es de vital importancia que los contenidos de los repositorios estén correctamente indizados en los motores de búsqueda.

Estas carencias encuentran un reflejo en la literatura científica, donde se observan muchos trabajos relacionados con el SEO para sedes web generales, pero muy pocos para SEO específico de repositorios digitales e institucionales. Algo similar ocurre con la disciplina de la cibermetría que, pese a haber generado una importante producción relacionada con el mundo académico en general y con la universidad en particular, los análisis cibernéricos aplicados a repositorios son todavía muy escasos.

Cibermetría aplicada al análisis de repositorios

El *Ranking Web of Repositories*¹¹, desarrollado en España por el *Cybermetrics Lab* en el seno del *Consejo Superior de Investigaciones Científicas* (CSIC), es probablemente el proyecto de mayor envergadura y peso en la aplicación de indicadores cibernéricos en el análisis de repositorios (Aguillo et al., 2010). Este proyecto posiciona los repositorios (tanto institucionales como temáticos) en función de cuatro indicadores web: tamaño (medido a través de *Google*), visibilidad (*MajesticSEO*¹² y *Ahrefs*¹³), número de ficheros ricos (*Google*) y tamaño académico (*Google Scholar*).

La elaboración de este producto ha permitido descubrir la inhabilidad a gran escala de los robots de los motores de búsqueda para recolectar indicadores web, debido a las barreras en el diseño de las bases de datos de los repositorios. Este problema concuerda parcialmente con los hallazgos ya comentados en los que el uso de esquemas de metadatos no recomendados es parte de la causa de la gran invisibilidad de los repositorios en *Google Scholar* (Arlitsch et al., 2012).

Igualmente, las carencias en la descripción de los metadatos influyen negativamente en la precisión de los indicadores cibernéricos, tal y como se demuestra en el análisis de *OpenAIRE* (Aguillo, 2011). Por otro lado, Orduña-Malea y Regazzi (2013),

descubren que las sintaxis utilizadas en los URLs de los repositorios de las universidades norteamericanas están influyendo negativamente en su visibilidad web.

Otros estudios se han centrado en el análisis de los orígenes de los hiperenlaces dirigidos hacia los repositorios institucionales. Por ejemplo, Sato y Isumura (2011) analizan los enlaces externos entrantes al repositorio institucional de la *Kyoto University*, descubriendo que una amplia mayoría de enlaces provenían de fuentes no académicas, tales como páginas personales y *Wikipedia*. Los enlaces provenientes desde páginas personales de académicos altamente citados hacia repositorios también ha sido estudiado recientemente (Mas-Bleda et. al, en prensa).

Por su parte, Smith (2011; 2012) analiza los enlaces externos recibidos por un conjunto de universidades australianas, para buscar su posible correlación con diversos indicadores bibliométricos relativos a la producción científica de las instituciones responsables de los repositorios.

En un primer análisis (Smith 2011), realizado a través de la herramienta *Bleeko*¹⁴, la autora identifica igualmente un amplio porcentaje de enlaces provenientes de *Wikipedia*, mientras que no encuentra correlación entre los indicadores bibliométricos y cibernéticos. Estos datos llevan a la autora a identificar dos posibles razones que justifiquen este hecho: por un lado la existencia de documentación en los repositorios que no guarda relación directa con la producción científica (trabajos de estudiantes, material gráfico, etc.), y por otro lado, la existencia de diversas motivaciones a la hora de crear un hiperenlace, distintas a las de la creación de una citación científica. Este análisis se expande posteriormente, utilizando *Google* en lugar de *Bleeko* (Smith, 2012), logrando resultados similares, lo que lleva a la autora finalmente a concluir que el valor de los repositorios institucionales parece centrarse en hacer la investigación accesible a la comunidad web general, más que a la comunidad científica en particular.

Este último argumento dirige de nuevo hacia las carencias en el tratamiento de los repositorios institucionales como meros sitios web (más allá de como baluartes del conocimiento científico), donde la consideración de ciertos indicadores y métricas pueden ayudar a monitorizar el uso real del repositorio y a mejorarlo para proporcionar mejores experiencias de uso a sus verdaderos usuarios.

En ese sentido, se destaca el trabajo de Scholze (2007), quien apunta diversos métodos para obtener datos de uso (*logs*) para los repositorios o Zuccala et. al (2007), quienes utilizan tanto análisis de enlaces como de ficheros *log* para analizar el impacto y uso de un repositorio institucional. Así mismo, en otro estudio (Zuccala et al., 2008) se utilizan técnicas avanzadas de análisis de enlaces como método para conocer potenciales usuarios o descubrir comunidades de usuarios ocultas, que resultan de interés para los gestores de los repositorios a nivel estratégico, al aportar información predecible y coherente.

En la actualidad, con la explosión de las llamadas *altmetrics*¹⁵, la disponibilidad de un amplio conjunto de indicadores de uso en los repositorios está comenzando a florecer, donde *Digital CSIC*¹⁶, en España, es un claro ejemplo de ello (Bernal, 2013), aunque su implantación en los repositorios institucionales es todavía escasa.

En el caso general de Latinoamérica, existen pocos trabajos que se centren en un análisis de los repositorios desde un punto de vista cibernético, pese a la amplia y compleja red de repositorios existentes en la actualidad. Se puede destacar el análisis cibernético básico del conjunto de repositorios de Argentina (Orduña-Malea, 2011).

En el caso más concreto de Costa Rica, es de gran valor el trabajo realizado por Córdoba (2011), quien realiza un exhaustivo trabajo de identificación y descripción de repositorios existentes, tanto académicos como gubernamentales.

2. Objetivos

En este trabajo se aplican diversos métodos cibernéticos para el análisis de un amplio conjunto de repositorios, con el propósito de mostrar que estas técnicas permiten obtener cierta información complementaria que ayuda a conocer mejor el impacto real de las plataformas de acceso abierto en la Web. Para ello, se plantean los siguientes objetivos específicos:

- a) A nivel de repositorio:
 - Conocer el nivel de interrelación de los repositorios, mediante un análisis de co-menciones de URL.
- b) A nivel de artículo o ítem:
 - Calcular las menciones de título y URL para los ítems de un repositorio dado (en este caso Kérwá).
 - Detectar las posibles malas prácticas que puedan afectar a la visibilidad de los ítems en los motores de búsqueda.

3. Metodología

En primer lugar se muestran los indicadores utilizados en el trabajo (aplicados a dos niveles: repositorio y artículo), y posteriormente se detallan los procesos de obtención de la muestra y su análisis estadístico.

3.1. Indicadores web

En la Tabla 1 se muestran los distintos indicadores utilizados en esta investigación, clasificados según su aplicación a la sede web de un repositorio, o bien a un artículo de un repositorio. Igualmente se indica la fuente utilizada para extraer la información así como una breve definición.

En la columna ‘Consulta’ se indica el método de obtención (directo a través de una consulta específica), en el motor de búsqueda seleccionado como fuente.

Tabla 1. Indicadores, fuentes y definición

NIVEL	CATEGORÍA	INDICADOR	FUENTE	DEFINICIÓN	CONSULTA
REPOSITORIO	MENCIÓN	URL mention	Google	Número de veces que el URL es mencionado.	“dominio.com” – site:dominio.com – inurl:dominio.com
		URL title	Google	Número de veces que el título de un artículo es mencionado.	“título” – site:dominio.com
ARTÍCULO	MENCIÓN	URL mention		Número de veces que el URL de un artículo es mencionado.	“dominio.com” – site:dominio.com

En el caso del análisis a nivel de repositorio, las medidas de mención se ciñen a la medición del número de veces que el URL es mencionado en entornos externos al repositorio.

En lo que se refiere a las medidas aplicadas a nivel de artículo, se obtienen dos indicadores: las menciones de título o “Title mention” (cantidad de veces que el título del ítem aparece en alguna de las páginas indizadas en *Google*) y las menciones de URL o “URL mention” (cantidad de veces que el URL que identifica al artículo aparece nombrado en algún documento). La razón del uso de estos indicadores viene motivada por la asunción de que si algo es enlazado o mencionado es porque tiene algún tipo de impacto o influencia (Kousha y Thelwall, 2006)

Estas dos medidas (Title mention y URL mentions) son estimativas del número de enlaces externos (Ortega et al., 2014), y son utilizadas a nivel de artículo pues las diversas fuentes de enlaces existentes (como *Open Site Explorer*, *Ahrefs* o *MajesticSEO*) no tienen cobertura suficiente para recursos individuales (en este caso los PDF de los diferentes artículos).

3.2. Selección y obtención de la muestra

Repositorios

En este caso, y con el propósito de obtener una muestra significativa, se procedió a seleccionar la totalidad de repositorios recogidos en el *Ranking Web of Repositories* (edición julio 2013) en el apartado “Latinoamérica”, y que constituyen un total de 137 repositorios. Los datos completos de los nombres de los repositorios, URL correspondiente y país se encuentran recogidos en el material complementario¹⁸.

Para cada uno de los 137 repositorios se recogió el URL y se comprobó su sintaxis para ser utilizada con precisión en el análisis cibernético (se recomienda el uso de subdominios dentro de la institución madre del repositorio).

En la Tabla 2 se muestran los URLs en los que se localizó algún tipo de incidencia, indicando asimismo si el URL fue finalmente medido y en qué condiciones. Como se puede observar, la mayoría de las incidencias vienen determinadas por el uso de subdirectorios en lugar de subdominios (pese a que a nivel técnico no existe diferencia entre ambos métodos, los subdirectorios presentan limitaciones en la precisión de los cálculos desde los motores de búsqueda).

Por otra parte, las redirecciones automáticas que se producen al acceder al recurso generan problemas en otros repositorios (aspectos que el webmaster debe resolver mediante una gestión adecuada del DNS).

Se debe indicar que para los URLs con subdirectorio, se probó a eliminar el subdirectorio para comprobar si existía una redirección. Si existía, se tomó en cuenta el URL madre; si no existía, daba error o se accedía a otro recurso, entonces los análisis se realizaron tomando en cuenta el subdirectorio. Por ejemplo:

<site:repositorio.utp.edu.co/dspace>

Tabla 2. Incidencias en la sintaxis de los URL de los repositorios

REPOSITORIO (URL)	INCIDENCIA	MEDIDO
intellectum.unisabana.edu.co:8080/jspui	No accesible	Sí
repositorio.utp.edu.co/dspace	Subdirectorio	Sí
uwispace.sta.uwi.edu/dspace	Subdirectorio (gestión DNS)	Sí
bdigital.ces.edu.co:8080/dspace	No localizado	No
repositorio.ufc.br	IP no localizada	No
bibliodigital.itcr.ac.cr/xmlui	Subdirectorio (gestión DNS; multidominio; no jerarquía con URL madre)	Sí
bibliodigital.itcr.ac.cr:8080/dspace	jerarquía con URL madre)	Sí
tesis.udea.edu.co/dspace	No localizado	No

repository.lasallista.edu.co/dspace	Subdirectorio (gestión DNS)	Sí
cedes.ufsc.br:8080/xmlui	Subdirectorio (URL madre es otro recurso)	No
ru.ffyl.unam.mx:8080/jspui	No localizado	No
repositorio.utfpr.edu.br/jspui/	Subdirectorio (gestión DNS)	Sí
repositorio.cti.gov.br/repositorio	Subdirectorio (URL madre otro recurso)	Sí
repositorio.ufma.br:8080/jspui	No localizado	No
repositorio.int.gov.br:8080/repositorio/	Subdirectorio (URL madre otro recurso)	No
campusp.uchile.cl:8080/dspace/	Subdirectorio (gestión DNS)	No
acervo.ufvjm.edu.br:8080/jspui/	Subdirectorio (URL madre otro recurso)	No
repositorio.ub.edu.ar:8080/xmlui	Gestión DNS	No
repositorio.ehtc.cu/jspui	Subdirectorio (gestión DNS; errores redirección)	Sí
biblio.colpos.mx:8080/jspui/	Subdirectorio (URL madre otro recurso)	No

Tras el proceso de filtrado, la muestra final quedó formada por un total de 127 URLs. Para cada uno de éstos se aplicó de manera manual el indicador de mención de título a nivel de repositorio, mostrado en la tabla 1.

Posteriormente, los datos obtenidos se volcaron a una hoja de cálculo para ser analizados estadísticamente mediante la aplicación *XLStat*.

El análisis a nivel de repositorio finaliza con un análisis de co-menciones. En este caso, y dada la complejidad de los cálculos, se consideraron únicamente los 50 repositorios con mayor tamaño en PDF medido por *Google*.

Para ello, una vez seleccionada la fuente, se creó una matriz de co-mención considerando todas las posibles combinaciones de menciones entre los 50 URLs, mediante una consulta como la siguiente:

<“lume.ufrgs.br” “teses.usp.br” –site:lume.ufrgs.br –site:teses.usp.br>

De esta forma, el buscador devuelve la cantidad de recursos en los que se han nombrado los dos URLs marcados (que corresponden con dos repositorios, excluyendo las propias menciones generadas en ambos sitios).

A partir de la aplicación *Webometric Analyst*¹⁹, se generó un fichero TXT con todas las posibles variantes de menciones (1.227). Cada variante se transformó en una consulta a *Google* (tal y como se especifica de forma genérica en la Tabla 1) y se procedió a realizar todas las consultas (una por variante) de forma automática a través de la API de *Bing* (debido a la imposibilidad de usar la API de *Google* en la actualidad para estos fines).

Con los datos obtenidos, se generó una matriz que se exportó a una hoja de cálculo para su procesamiento estadístico, realizando un análisis MDS a través de la suite *XLStat*. De forma paralela, se extrajeron estadísticas de indicadores topológicos de la red de menciones obtenida (*Degree Binary*, *Degree weighted* y Densidad de la red), mediante las facilidades de *Webometric Analyst*.

Finalmente, se generaron dos mapas de co-menciones:

- Mediante el *Webometric Analyst Network Graph Tool*.
- Mediante la aplicación *Gephi* (mostrando sólo el 75% de las menciones obtenidas para simplificar el mapa).

Artículos

Para el análisis a nivel de artículo se eligió un repositorio concreto: *Kérwá*, el repositorio institucional de la *Universidad de Costa Rica* (kerwa.ucr.ac.cr).

La razón para escoger este repositorio vino determinada el deseo de escoger un repositorio representativo de Costa Rica (de entre los existentes en el *Ranking Web of Repositories*) cuyo tamaño fuera asequible para realizar un análisis en profundidad. Los

557 recursos localizados en el momento del análisis suponían una cantidad razonable para poder plantearse el análisis completo de la colección, algo más complicado en el caso de repositorios con una cantidad de ítems mayor.

Una vez seleccionado el repositorio, se procedió a listar todos los ítems depositados en el mismo, y a exportar a una hoja de cálculo tanto el título del recurso como el URL de acceso al mismo. A continuación, se calculó el número de menciones externas tanto al título como al URL (tanto de la URI del recurso como del fichero PDF final de descarga) para cada ítem a través de *Google*:

Mención de título:

< "Generalidades de serpientes venenosas en Costa Rica" –site:ucr.ac.cr >

Mención de URL:

< "hdl.handle.net/10669/867" –site:ucr.ac.cr >

< "kerwa.ucr.ac.cr/handle/10669/867" –site:ucr.ac.cr >

< "kerwa.ucr.ac.cr/bitstream/handle/10669/867" –site:ucr.ac.cr >

Los tres URLs medidos vienen determinados por la existencia de diferentes formas de acceso a cada uno de los ítems (debido al propio mecanismo de funcionamiento del software). En la figura 1 se observa la identificación de los distintos URLs considerados (el URL del navegador, URI oficial y el URL del recurso final), distintos todos ellos, por lo que precisan de análisis independientes.

The screenshot shows a DSpace repository page. At the top, the browser address bar displays `kerwa.ucr.ac.cr/handle/10669/580`. The page header includes the 'Kerwa' logo and the text 'REPOSITORIO INSTITUCIONAL DE LA UNIVERSIDAD DE COSTA RICA'. The main content area features the title 'Formación de la clase media en Costa Rica. Economía, sociabilidades y discursos políticos (1890-1950)' and the author 'García, George I.'. The abstract and description are provided. A table titled 'Files in this item' lists a file named 'tesis_completa.pdf' with a size of 2.832Mb and a PDF format. The URI 'http://hdl.handle.net/10669/580' is highlighted in the description field, and the file name 'tesis_completa.pdf' is highlighted in the table.

Figura 1. Ficha de un artículo depositado en el repositorio *Kerwá*

En la figura 2 se muestra, a modo de ejemplo, el proceso de cálculo del número de menciones de URL para una de las variantes, a través de *Google*.

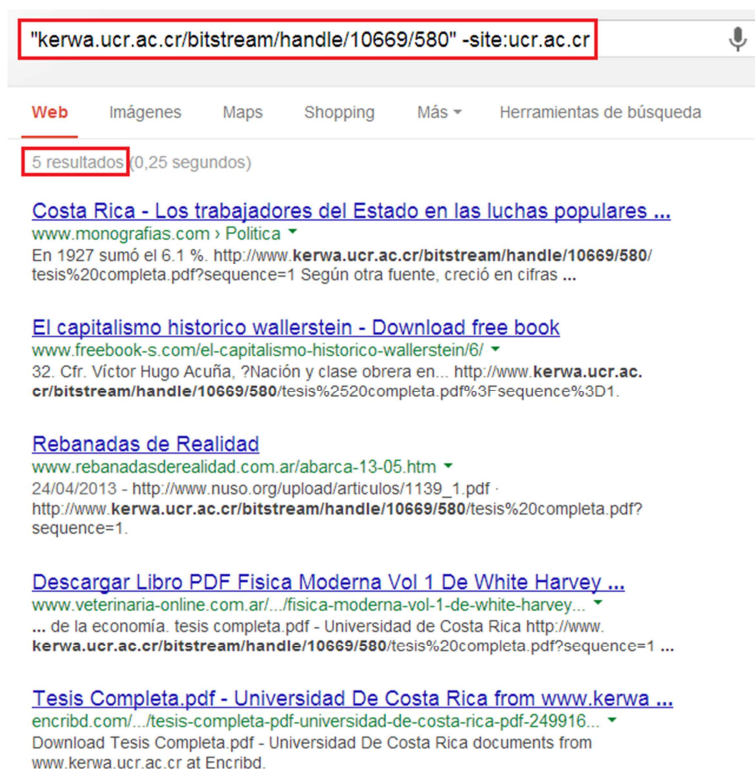


Figura 2. Ejemplo de consulta de menciones de URL para un artículo en Google
Google and Google logo son marcas registradas de Google Inc.

4. Resultados

4.1. Análisis de los repositorios latinoamericanos

4.1.1. Distribución geográfica

En la tabla 3 se muestra la distribución por país de los 137 repositorios analizados, donde se puede observar el predominio de Brasil (37), Colombia (21), Argentina (18) y Ecuador (17). Adicionalmente, se incluye el número de repositorios de esos países situados en el top 50 del Ranking Web, donde sobresale Brasil (12) y Ecuador (7).

Tabla 3. Distribución de países de la muestra de repositorios

PAÍS	N	TOP 50
Brasil	37	12
Colombia	21	7
Argentina	18	7
Ecuador	17	10
México	12	3
Chile	8	4
Venezuela	8	3
Perú	5	2
Costa Rica	3	0
Cuba	3	0
El Salvador	2	1
Jamaica	2	1
Puerto Rico	1	0
TOTAL	137	50

En el caso de Chile, se debe tener en cuenta que un repositorio (*Repositorio Digital CONICYT*), dispone de dos URLs (<dspace.conicyt.cl/ri20> y <dspace2.conicyt.cl>) que, a efectos de este trabajo, se han tratado de forma independiente.

4.1.2. Análisis de co-menciones

Con el objetivo de conocer el grado de relación entre los repositorios de la muestra, se aplica un análisis de co-menciones, en el que se consideran únicamente los 50 repositorios con mayor número de ítems PDF en Google (1.225 combinaciones), pues es el indicador que mejor correlaciona con el número total de ítems (es decir, se recogen los repositorios con las colecciones mejor indizadas en *Google*).

En la Tabla 4 se muestran las combinaciones que aparecen con mayor frecuencia. Los datos muestran un elevado grado de co-relación; sólo 8 parejas de URLs aparecen sin ninguna co-mención, mientras que 39 superan las 50 y 4 superan las 100, un valor que, dada la muestra, se estima elevado. Esto se refleja en la alta densidad de la red calculada ($D= 0.9935$), y en los valores de centralidad calculados (Tabla 5).

Tabla 4. Repositorios con mayor grado de co-mención (n=50)

PAREJA DE URLs	CO-MENCIONES
lume.ufrgs.br teses.usp.br	188
teses.usp.br dspace.c3sl.ufpr.br	125
dspace.espol.edu.ec dspace.ups.edu.ec	111
dspace.espol.edu.ec bibdigital.epn.edu.ec	108
dspace.espol.edu.ec dspace.esepoch.edu.ec	95
dspace.ups.edu.ec bibdigital.epn.edu.ec	89
saber.ula.ve bdigital.unal.edu.co	85
dspace.espol.edu.ec bdigital.unal.edu.co	84
dspace.espol.edu.ec repositorio.espe.edu.ec	80
dspace.ups.edu.ec repositorio.espe.edu.ec	78
dspace.ups.edu.ec dspace.esepoch.edu.ec	75
teses.usp.br maxwell.lambda.ele.puc-rio.br	70
bdigital.unal.edu.co repository.urosario.edu.co	70
repositorio.espe.edu.ec bibdigital.epn.edu.ec	70
bibdigital.epn.edu.ec dspace.esepoch.edu.ec	70
bdigital.unal.edu.co dspace.ups.edu.ec	66
saber.ula.ve tesis.ula.ve	63
sedici.unlp.edu.ar memoria.fahce.unlp.edu.ar	63
dspace.espol.edu.ec repositorio.utn.edu.ec	62
lume.ufrgs.br dspace.c3sl.ufpr.br	61

Tabla 5. Nodos con mayor Degree (Binary y weighted)

Repositorio	Degree (Weighted)	Degree Binary
dspace.espol.edu.ec	89,000	49
dspace.ups.edu.ec	77,000	49
bdigital.unal.edu.co	73,000	49
saber.ula.ve	69,000	49
teses.usp.br	68,000	49
bibdigital.epn.edu.ec	67,000	49
dspace.esepoch.edu.ec	65,000	49
repositorio.espe.edu.ec	64,000	49
lume.ufrgs.br	60,000	49
sedici.unlp.edu.ar	57,000	49

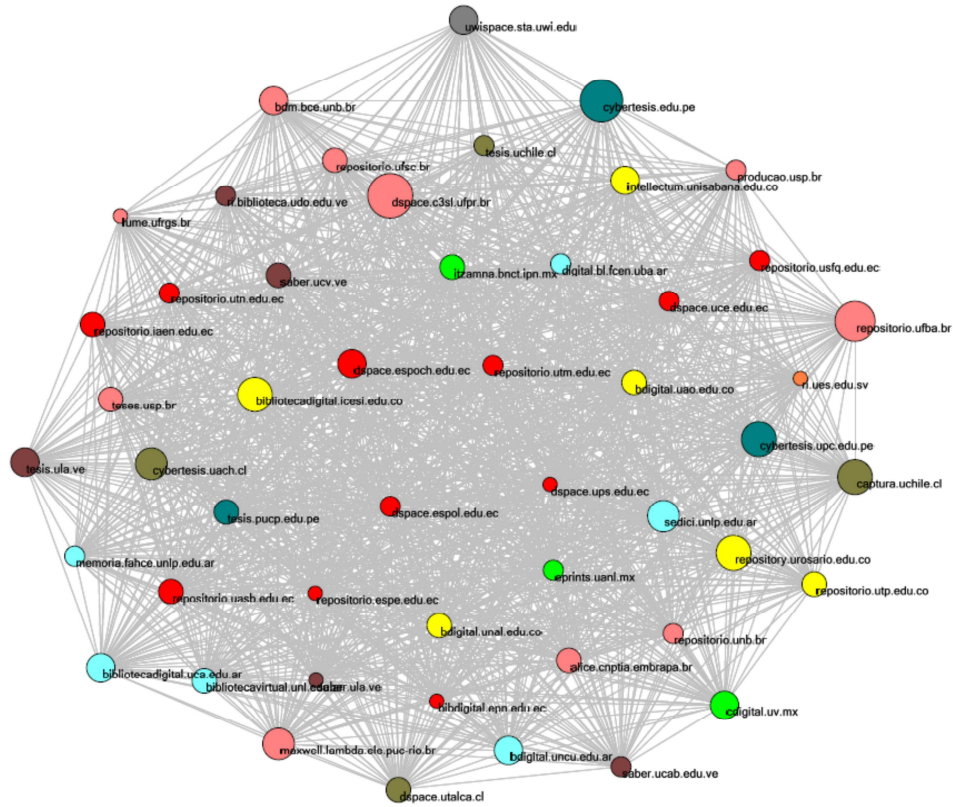


Figura 4. Mapa de co-menciones (n=50; *Webometric Analyst*)

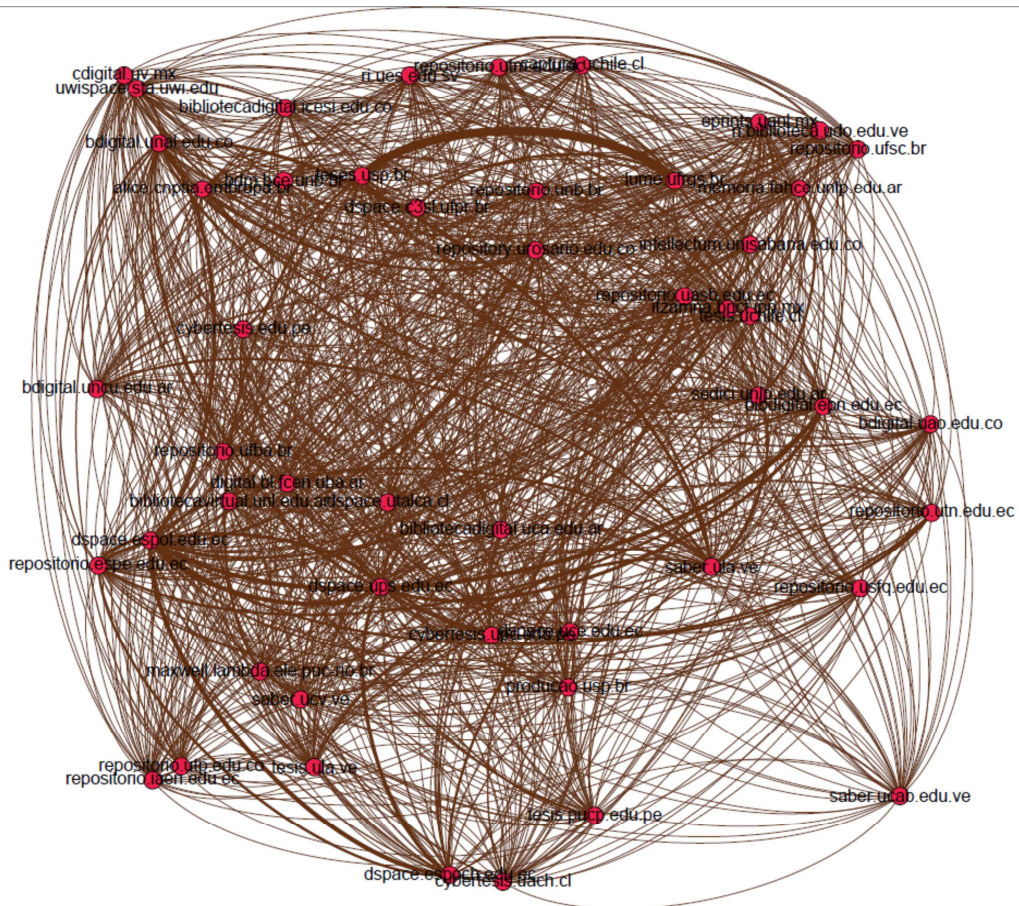


Figura 5. Mapa de co-menciones (n=50; *Gephi*, proporción de previsualización: 75%)

4.2. Análisis del repositorio institucional de la *Universidad de Costa Rica (Kérwá)*

Para finalizar, se realiza un análisis del repositorio *Kérwá*, analizando la visibilidad web de su colección. Para ello, como se ha indicado anteriormente, se calcula menciones de título y URL para cada uno de los ítems depositados en el repositorio.

En la tabla 6 se muestran los títulos que han logrado un mayor número de hits en *Google*.

Los valores logrados son bastante desiguales, por un lado se observan datos muy elevados (por ejemplo, “Anemia de células falciformes” obtiene más de un millón e hits), mientras que 16 títulos no logran una sola mención externa, y en otros 397 no se sobrepasa la decena de resultados, por lo que la distribución es muy desigual.

Uno de las posibles razones es el ruido que introduce este indicador (similar al de las menciones de URL, aplicado anteriormente). Por ello, se puede observar cómo los títulos con mayor número de menciones son genéricos y comunes (es decir, no tienen por qué relacionarse necesariamente con el título del artículo). Ejemplos de ello son “Ensayos políticos” o “Programación en HTML”, por lo que este indicador sólo puede usarse de manera indicativa y con mucha cautela.

Tabla 6. Menciones de título con mayor número de hits

TÍTULO	HITS
Anemia de células falciformes	1.460.000
La técnica en la economía y la investigación	436.000
La cultura de información	95.000
Los fenómenos de El Niño y La Niña	89.900
Ética, política y sociedad	80.600
Música para orquesta de cuerda	48.500
Ensayos políticos	47.500
Programación con HTML	45.900
Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing	34.200
Población y Salud en Mesoamérica	32.900
Colección Identidad Cultural	28.100
Economía, sociedad y ambiente	27.400
II Seminario "Estado de la Investigación en el Caribe de Costa Rica"	26.800
Instituto Clodomiro Picado	25.400
Banco de germoplasma de Pejibaye.	18.100
La alborada del capitalismo agrario en Costa Rica	16.400
Envenomations by Bothrops and Crotalus Snakes Induce the Release of Mitochondrial Alarmins	16.300
A Lys49-PLA2 myotoxin of Bothrops asper triggers a rapid death of macrophages that involves autocrine purinergic receptor signaling	11.800
Investigaciones de mercado para la calidad	11.200
Suero antiofidico polivalente	11.200

En cuanto a las menciones de URL de los diferentes ítems, los valores logrados han sido muy bajos. En la tabla 7 se muestra, para las tres tipologías de URLs analizadas, los valores acumulados obtenidos. La columna “Hits total” muestra la cantidad de resultados para cada variante de URL mostrada en la tabla, mientras que la columna “Hits Suma” muestra el sumatorio de enlaces para cada ítem (añadiendo el código correspondiente como subdirectorío a la URL madre).

Tabla 7. Menciones de URL para los ítems del repositorio

URL	Hits Total	Hits Suma
"hdl.handle.net/10669" -site:ucr.ac.cr	2.260	492
"kerwa.ucr.ac.cr/bitstream/handle/10669" -site:ucr.ac.cr	1.550	211
"kerwa.ucr.ac.cr/handle/10669" -site:ucr.ac.cr	405	40

Por un lado, los datos indican un mayor uso de la variante <hdl.handle.net/10669>. Por otro lado, las excesivas diferencias de ambos cálculos (que deberían haber proporcionado resultados similares), se aducen a limitaciones del buscador en el caso de las consultas al URL general ("Hits Total").

5. Conclusiones

Las principales conclusiones obtenidas de este trabajo son las siguientes:

Análisis de co-menciones

El análisis de las co-menciones entre repositorios indica una densidad de red muy elevada ($D=0.9935$). Los repositorios aparecen nombrados conjuntamente en multitud de ocasiones, lo que muestra una alta cohesión entre ellos de manera externa. No obstante, estos datos quedan marcados por la fuerte influencia de los repositorios brasileños y ecuatorianos en la muestra.

Análisis de ítems

El estudio a nivel de ítem ha demostrado ser complicado debido al ruido documental de algunos indicadores (menciones de título) y a los valores excesivamente bajos en el caso de los enlaces recibidos (menciones de URL). Estas medidas, provocadas en gran medida por los protocolos de generación de URLs dentro del repositorio (*handle*) deberían ser complementadas en el futuro con datos de uso del repositorio a nivel de artículo (especialmente visitas y descargas).

6. Notas

1. <http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung>
2. <http://www.budapestopenaccessinitiative.org>
3. <http://arxiv.org>
4. <http://repec.org>
5. <http://dspace.mit.edu>
6. http://europa.eu/rapid/press-release_IP-13-786_es.htm
7. <http://oaseminar.fecyt.es/Publico/AlhambraDeclaration/index.aspx>
8. <http://creativecommons.org>
9. <http://sciencecommons.org/about>
10. E-Expectations research reports: <http://omniupdate.com/resources/research.html>
11. <http://repositories.webometrics.info>
12. <http://www.majesticseo.com>
13. <http://ahrefs.com>
14. <http://blekko.com>
15. <http://altmetrics.org/manifesto>
16. <http://digital.csic.es/estadisticas/estadisticasUso.jsp>
17. <http://www.opensiteexplorer.org/>
18. Material complementario: <http://digibug.ugr.es/handle/10481/28516>
19. <http://lexiurl.wlv.ac.uk>

7. Referencias

- Aguillo, I. F. (2009). "Measuring the institutions' footprint in the web". *Library Hi Tech*, v. 27, n. 4, pp. 540-556.
- Aguillo, I. F. (2011). "Building web indicators for the EU OA repository". *Workshop on new research lines in informetrics*, Madrid (España), 16 mayo.
- Aguillo, I. F.; Ortega, J. L.; Fernandez, M.; Utrilla, A. M. (2010). "Indicators for a webometric ranking of open Access repositories". *Scientometrics*, v. 82, n. 3, pp. 477-486.
- Arlitsch, K.; O'Brian, P. S. (2012). "Invisible institutional repositories: addressing the low indexing ratios of IRs in Google". *Library Hi Tech*, v. 30, n. 1, pp. 60-81.
- Bernal, I. (2013). "Open access and the changing landscape of research impact indicators: new roles for repositories". *Publications*, v. 1, pp. 56-77.
- Córdoba, S. (2011). "Los repositorios institucionales y de acceso abierto en Costa Rica". *E-Colabora: Revista de ciencia, educación, innovación y cultura apoyadas por Redes de Tecnología Avanzada*, v. 1, n. 2, pp. 86-100.
- De-Castro, P. (2014). "Desarrollo de servicios para repositorios en el Reino Unido: el proyecto UK RepositoryNet+". *Anuario ThinkEPI 2014* [en prensa].
- De Rosa, C.; OCLC (2005). Perceptions of libraries and information resources; a report to the OCLC membership. *OCLC Online Computer Library Center*, Dublin, OH.
- DeRidder, J. L. (2008). "Googlizing a digital library". *The Code4Lib Journal*, n. 2. <http://journal.code4lib.org/articles/43>
- Kousha, K., & Thelwall, M. (2006). Motivations for URL citations to open access library and information science articles. *Scientometrics*, 68(3), 501-517.
- Mas-Bleda, A. Thelwall, M., Kousha, K.; Aguillo, I. (en prensa). "Successful researchers publicizing research online: an outlook analysis of European highly cited scientists' personal websites". *Journal of Documentation*.
- Orduña-Malea, E. (2011). "Visibilidad de los repositorios institucionales argentinos en la Web: indicadores y buenas prácticas". En: *TIEB2: Segundo Taller de Indicadores de Evaluación de Bibliotecas*. La Plata (Argentina), 27 y 28 de junio, 27 p.
- Orduña-Malea (2013). "Aggregation of the web performance of internal university units as a method of quantitative analysis of a university system: the case of Spain". *Journal of the American Society for Information Science and Technology*, v. 64, n.10, pp. 2100-2114.
- Orduña-Malea, E.; Regazzi, J. J. (2013). "US academic libraries: understanding their web presence and their relationship with economic indicators". *Scientometrics* [Online first].
- Ortega, J. L., Orduña-Malea, E.; Aguillo, I. F. (2014). "Influence of language and file type on the web visibility of top European universities". *Aslib Proceedings*, v. 66, n. 1, pp. 96-116.
- Sánchez García, S.; Melero, R. (2006). "La Denominación y el contenido de los repositorios institucionales en acceso abierto: Base teórica para la 'Ruta verde'". [No publicado]. http://eprints.rclis.org/7613/1/Denominaci%C3%B3n_contenido_OA.pdf
- Sato, S.; Isumura, H. (2011). "How do people use open access papers in non-academic activities? A link analysis of papers deposited in institutional repositories". *Library, Information and Media Studies*, v. 9, n. 1, pp. 51-64.

- Scholze, F. (2007). "Measuring research impact in an open access environment". *Liber Quarterly: The Journal of European Research Libraries*, v. 17, n. 1-4, pp. 220-232.
- Smith, A. G. (2011). "Wikipedia and institutional repositories: An academic symbiosis?". En *Proceedings of the ISSI 2011 Conference. Durban (South Africa)*, 4-7 Julio, pp. 794-800.
- Smith, A. G. (2012). "Webometric evaluation of institutional repositories". En: *Proceedings of the 8th International Conference on Webometrics Informetrics and Scientometrics & 13th Collnet Meeting. Seoul (Korea)*, pp. 722-729.
- Smith, A.G. (2013). "Web Based Impact Measures for Institutional Repositories". En *Proceedings of the ISSI 2013 conference. Viena (Austria)*, pp. 1806-1816.
- Zuccala, A.; Thelwall, M.; Oppenheim, C.; Dhiensa, R. (2007). "Web intelligence analyses of digital libraries: A case study of the national electronic library for health (NeLH)". *Journal of Documentation*, v. 63, n. 4, pp. 558-589.
- Zuccala, A.; Oppenheim, C.; Dhiensa, R. (2008). "Managing and evaluating digital repositories". *Information research*, v. 13, n. 1.
<http://informationr.net/ir/13-1/paper333.html>