



Diseño e implementación de una *Master List* de revistas científicas a partir de Ulrich's, Web of Science, y Scopus.

Alberto Martín Martín

Trabajo de Fin de Máster
Tutor: Emilio Delgado López-Cózar
Máster en Información y Comunicación Científica

Índice

| | |
|---|----|
| Declaración de originalidad | 3 |
| 1. Introducción | 4 |
| 1.1. Justificación | 4 |
| 1.2. Estado de la cuestión | 7 |
| 2. Objetivos | 10 |
| 3. Material y métodos | 11 |
| 3.1. Ulrich's International Periodicals Directory | 11 |
| Descarga | 11 |
| Procesamiento | 12 |
| 3.2. WoS Master Journal Lists / Journal Citation Reports. | 15 |
| Descarga | 15 |
| Procesamiento | 16 |
| 3.3. Scopus Title List / SCImago Journal Rank..... | 17 |
| Descarga | 17 |
| Procesamiento | 18 |
| 3.4. Integración de las tres bases de datos..... | 19 |
| 4. Resultados | 22 |
| 4.1. Resultados generales. Activas/Inactivas, y control del ISSN y de títulos..... | 22 |
| 4.2. Análisis de las clasificaciones temáticas..... | 25 |
| 4.3. Estudio del solapamiento | 27 |
| 4.4. Análisis de la cobertura de las bases de datos: geográfica, lingüística, editorial y temática | 29 |
| Cobertura geográfica | 29 |
| Cobertura lingüística | 30 |
| Cobertura editorial | 32 |
| Cobertura temática | 34 |
| 5. Conclusiones..... | 35 |
| 6. Referencias bibliográficas..... | 37 |
| Anexo 1..... | 40 |
| Anexo 2 | 41 |

Declaración de originalidad

El alumno, D. Alberto Martín Martín, garantiza, al firmar este Trabajo Fin de Máster, que este trabajo de investigación ha sido realizado respetando los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, viernes, 31 de enero de 2014

El alumno

Fdo.: Alberto Martín Martín

1. Introducción

1.1. Justificación

Las revistas científicas, desde su creación a finales del siglo XVII, se han convertido en el principal medio de comunicación formal en la gran mayoría de las disciplinas científicas. Estas revistas cumplen cinco funciones básicas (Delgado López-Cózar & Ruiz-Pérez, 2009; Delgado López-Cózar *et al.*, 2006):

1. Selección de los mejores originales que reciben.
2. Control de calidad de la investigación.
3. Mejora de la legibilidad de los textos.
4. Otorgamiento de crédito, gracias a la reputación adquirida por los títulos largamente establecidos.
5. Archivo del conocimiento.

Durante el siglo XX, y especialmente tras la Segunda Guerra Mundial, el número de publicaciones periódicas alcanzó un ritmo de crecimiento vertiginoso. En el caso de las revistas científicas, este fenómeno fue detectado en los años sesenta por Price, que postuló que el número de revistas científicas estaba creciendo de forma exponencial. Este hecho, combinado con las dificultades que conlleva el control de tantas publicaciones, ha dado lugar a la aparición de diversos directorios dedicados al control de las publicaciones periódicas. Los más importantes a nivel mundial son el registro internacional del ISSN, y el directorio Ulrich's (Delgado López-Cózar, 1996):

- **El registro internacional del ISSN:** gestionado por el Centro Internacional para el Registro de Publicaciones Seriadas (CIEPS), también llamado Centro Internacional del ISSN. Es la mayor fuente de información en el mundo para la identificación de publicaciones seriadas. Su base de datos, la cual se puede acceder mediante suscripción, contiene alrededor de 1,7 millones de ISSN.
- **Ulrich's International Periodicals Directory:** desde que comenzó a ser editado en 1932 en forma impresa por Carolyn Ulrich, por entonces jefa de la sección de publicaciones periódicas de la Biblioteca Pública de Nueva York, este directorio se ha convertido en una de las fuentes de información más exhaustivas en todo lo referente a publicaciones periódicas. Actualmente contiene alrededor de 700.000 registros. El acceso a este producto también es por suscripción, y ofrece una gran cantidad de información bibliográfica muy útil sobre las publicaciones, como son la lista de bases de datos en las que está indizada cada revista, si es de acceso abierto, o si realiza revisión por pares.

Debido precisamente a su status como depósito de los resultados de las investigaciones que realizan los científicos (en forma de artículos), y también como

herramientas de evaluación del rendimiento de los propios investigadores, las revistas científicas se han convertido en objeto de estudio común en el campo de la bibliometría, disciplina que estudia y analiza la actividad científica.

Desde que el *Institute for Scientific Information* (ISI), con Eugene Garfield a la cabeza, publicó la primera edición de los *Journal Citation Reports* en 1975 (Garfield, 2007), han seguido apareciendo una gran cantidad de productos que analizan las revistas científicas de una u otra forma.

A nivel internacional destacan:

- los ya citados ***Journal Citation Reports***, en sus dos ediciones: Science y Social Sciences, nutridos respectivamente por el *Science Citation Index*, y el *Social Sciences Citation Index* (publicados por Thomson Reuters desde 1992). Es aquí donde podemos encontrar el famoso *Impact Factor*, indicador que pone en relación el número de trabajos que publica una revista con el número de citas que reciben estos trabajos.
- El indicador ***SCImago Journal Rank*** (SJR), desarrollado por el grupo de investigación SCImago (del CSIC) a partir de los datos de citación de la base de datos Scopus, publicada por la editorial Elsevier. Este indicador, al igual que el *Impact Factor*, también tiene en cuenta el número de trabajos publicados por una revista, y el número de citas recibidas, pero además, imitando al algoritmo PageRank de Google, también tiene en cuenta el prestigio de la fuente citante. Esto lo hace asignando diferentes pesos a cada cita en función de su procedencia, mediante un proceso iterativo (Guerrero-Bote & Moya-Anegón, 2012).
- Recientemente, Google también ha entrado al campo de la edición de productos bibliométricos. Entre ellos, a nivel de revistas científicas tenemos ***Google Scholar Metrics*** (GSM), lanzado en 2012, que toma como base, como su propio nombre indica, los datos recogidos en *Google Scholar* (Delgado López-Cozar et al., 2013). *Google Scholar Metrics* ofrece un ranking de revistas según su índice h (Delgado López-Cózar & Cabezas-Clavijo, 2012). El índice h, propuesto por Jorge Hirsch, de la Universidad de California, también es un indicador basado en la citación. Un científico (o en este caso, una revista) tiene un índice h si ha publicado h trabajos con al menos h citas cada uno.
- El ***European Reference Index for the Humanities*** (ERIH), desarrollado por la European Science Foundation, es otro producto que clasifica las revistas en varias categorías, en función de unos determinados criterios de calidad. Este producto está sirviendo para suplir las deficiencias de cobertura de las principales índices internacionales en las áreas de Humanidades y Ciencias Sociales.
- ***Latindex***: iniciado por la Universidad Nacional Autónoma de México en 1995, Latindex ofrece, entre otros productos, un catálogo de publicaciones periódicas

de investigación científica, técnico-profesionales, y de divulgación científica que se editan en los países de América Latina, el Caribe, España, y Portugal. Todas las revistas de este catálogo han sido evaluadas en función de unos criterios formales, en los que se encuentran parámetros relativos a la presentación de la revista, la gestión y la política editorial, y las características de los contenidos.

A nivel nacional, en España también han aparecido una serie de productos de carácter bibliométrico. En muchos casos, estos productos también han estado inspirados por la necesidad de cubrir las lagunas que tienen las bases de datos internacionales en las áreas de Ciencias Sociales y Humanidades:

- **RESH** (Revistas Españolas de Ciencias Sociales y Humanidades), es una herramienta desarrollada por el Grupo de Investigación de Evaluación de Publicaciones Científicas (EPUC) del CSIC, y por el Grupo de Investigación de la Ciencia y la Comunicación Científica (EC3) de la Universidad de Granada. Esta herramienta por un lado recopila la lista de criterios de calidad que cumple cada revista según la CNEAI, la ANECA, y *Latindex*, y por otro lado, ofrece otros dos indicadores: el de difusión (según la presencia de la revista en bases de datos nacionales e internacionales), y el de “opinión de los expertos” (calculado a partir de los resultados de una encuesta a profesores e investigadores).
- **DICE** (Difusión y Calidad Editorial de las Revistas Españolas de Humanidades y Ciencias Sociales), también desarrollado por EPUC. Evalúa a las revistas según una serie de criterios formales relacionados con las políticas de selección de originales para publicación, la presencia en bases de datos, apertura exterior de los autores, e internacionalidad de los autores. En enero de 2013 EPUC difundió un comunicado en el que se anunciaba el cese de la actualización de este producto debido al corte de su financiación por los recortes que afectan a la Administración.
- **MIAR** (Matriz de información para la evaluación de revistas), desarrollado en la Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona. Su objeto de estudio son las revistas científicas de Ciencias Sociales y Humanas a nivel internacional, con el objetivo de “situar y contrastar las revistas españolas en el plano internacional”. En este producto encontramos el indicador ICDS (Índice Compuesto de Difusión Secundaria), cuyo valor depende de la presencia de la revista en determinadas bases de datos, según su ámbito científico.
- **IN-RECS, IN-RECH e IN-RECI**: Índice de impacto de las revistas españolas de Ciencias Sociales, Ciencias Humanas, y Ciencias Jurídicas (respectivamente) desarrollados por el grupo EC3 (Delgado López-Cózar *et al.*, 2005). Este producto emula el *Impact Factor* de Thomson Reuters, con la diferencia de que utiliza una ventana de citación de tres años, en vez de dos. Además, hasta su edición de 2009, ofrecía una información muy detallada acerca de la evolución temporal y los

patrones de citación entre las revistas, lo cual, entre otras cosas, facilitaba la detección de prácticas sospechosas. Estos productos también han sido recientemente discontinuados por falta de financiación.

- **CIRC** (Clasificación Integrada de Revistas Científicas). Gestionado también por miembros del grupo EC3 (Torres-Salinas *et al.*, 2010). Esta herramienta tiene como objetivo “la construcción de una clasificación de revistas científicas de Ciencias Sociales y Humanas en función de su calidad integrando los productos de evaluación existentes considerados positivamente por las diferentes agencias de evaluación nacionales como CNEAI y ANECA”.

Debido a la gran cantidad de productos bibliométricos que indizan y evalúan las revistas científicas, para elaborar una imagen que refleje todas las facetas de una misma revista en un momento determinado, o su evolución a lo largo del tiempo, es necesario consultar estos productos uno por uno. Además, muchos de estos productos son tenidos en cuenta por las agencias públicas de evaluación científica españolas (como CNEAI y ANECA) en sus convocatorias de evaluación.

En este contexto, en el que con frecuencia se encuentran los investigadores dedicados a la bibliometría, surge una necesidad muy clara: una herramienta desde la que se puedan consultar todos los datos de los productos anteriormente reseñados desde un único sitio. Esto conlleva la creación de una base de datos en la que se integren los diversos productos de los que se han hablado. En esta base de datos habrá que prestar especial atención a la normalización de las revistas entre los diferentes productos, con el objetivo de unir los registros de una misma revista en cada uno de los productos integrados. Además, debería estar diseñada para poder ser actualizada regularmente. En definitiva, estamos hablando de una *Master List* de revistas científicas.

1.2. Estado de la cuestión

Existe una abundante literatura científica dedicada al estudio comparativo de las bases de datos e índices bibliométricos a fin de detectar sus similitudes y diferencias. A comienzos del siglo XXI, Braun *et al.* (2000) publicaron un estudio sobre la cobertura temática del Science Citation Index (SCI), comparándolo con el directorio Ulrich's. Según este estudio, contrariamente a las creencias generales, en la gran mayoría de los resultados la cobertura del SCI demostró ser bastante equilibrada comparada con la cobertura de Ulrich y no se encontraron sesgos que dieran más peso a las ciencias biomédicas entre todas las disciplinas, y a Estados Unidos entre los países. En Archimbault *et al.* (2006) también se hace una comparación con Ulrich's, y en este caso además del SCI, se incluye en el análisis el Social Sciences Citation Index (SSCI), y el Arts & Humanities Citation Index (A&HCI). En este estudio se compara la cobertura entre estos productos a nivel del idioma, detectando una sobrerrepresentación del inglés en las bases de datos de Thomson

Reuters. Por su parte, en Moya-Anegón *et al.* (2007) se realiza un estudio comparativo de la cobertura a nivel temático, geográfico (por países y por idiomas), y por editoriales, entre Scopus, y el directorio Ulrich's. A nivel temático se concluyó que la cobertura de Scopus era bastante equilibrada comparada con la de Ulrich's, excepto en Arte y Humanidades. A nivel geográfico, los países en los que se editaban más revistas eran Estados Unidos y Reino Unido, con mucha más representación que en Ulrich's. Aproximadamente el 74% de las revistas de Ulrich's estaban escritas en inglés, mientras que este porcentaje subía hasta un 85% en Scopus. En cuanto a las editoriales, en las dos bases de datos las más importantes eran Elsevier y Blackwell, aunque ésta última estaba mejor representada en Ulrich's.

Tampoco faltan estudios que comparen las principales bases de datos de citas del mercado: Web of Science y Scopus. Así, Goodman & Deis (2005, 2007), LaGuardia (2005), y Fingerman (2006) realizaron cada uno sus respectivas comparativas entre los productos poco después de que apareciera Scopus en 2004. Estas reseñas se centran sobre todo en las funcionalidades que estaban disponibles en las interfaces de los productos en aquella época y sus precios, aunque también realizaban una somera comparativa de la cobertura temática, y en función de esto se planteaban la disyuntiva de seleccionar uno u otro producto, para adquirirlo y ofrecerlo a los usuarios de las bibliotecas académicas. En Gavel & Iselid (2008) se hace un completo estudio del solapamiento de revistas entre las dos bases de datos WoS y Scopus, incluyendo también algunas bases de datos especializadas como Medline y Embase, y utilizando además el directorio Ulrich's de forma complementaria para enriquecer la información bibliográfica de los registros. En este estudio también se pone de manifiesto la mayor cobertura de revistas que hay en Scopus, excepto en el campo de las humanidades, donde existían un gran número de revistas que solo se encontraban en WoS. En Archimbault *et al.* (2009) se comparan de nuevo Web of Science y Scopus, esta vez haciendo un estudio del número de artículos y citas por países, concluyendo que hay una altísima correlación entre los resultados de los dos productos. Por último, en Leydesdorff *et al.* (2010) se comparan las matrices de citación revista-revista que se consiguen tanto con los datos de Web of Science como con los datos de Scopus. Debido a la mayor cobertura de revistas en Scopus, se esperaba que los mapas de citación que se pueden generar con estos datos fueran más grandes que los que se generan con datos de WoS, pero esto en ocasiones no sucede porque WoS contiene un historial de citación mayor en las citas más antiguas (más de 10 años).

Además de comparaciones entre las bases de datos de la Web of Science y Scopus, también existen comparaciones entre los indicadores bibliométricos que se calculan a partir de los datos de uno y otro producto. Falagas *et al.* (2008) es un ejemplo de ello. En este estudio se compara el factor de impacto calculado en los Journal Citation Reports (JCR) de Thomson Reuters, con el SCImago Journal Rank (SJR) desarrollado por el grupo de investigación SCImago con datos de Scopus, y

concluye que éste último es una alternativa seria al tradicional factor de impacto debido a que es libremente accesible, tiene una mayor cobertura de revistas, y al hecho de que tiene en cuenta el prestigio de la fuente citante. En López-Illescas *et al.* (2008) se hace un estudio comparativo restringido al campo de la oncología, y pone de manifiesto que las revistas de oncología que aparecen en WoS se pueden considerar un subconjunto de las que aparecen Scopus (ya que en esta última, el número de revistas sobre oncología es mayor, y se engloban la gran mayoría de las revistas de WoS), y que el factor de impacto calculado en los JCR tiene una alta correlación con el SJR.

En los últimos años ha aparecido un nuevo jugador en el campo de las bases de datos académicas de citas y de los productos de evaluación científica: Google Scholar (GS). Esto no ha pasado desapercibido para la comunidad científica, y se ha traducido en un gran número de estudios que comparan las bases de datos académicas de citas (y también especializadas), con Google Scholar. En Falagas *et al.*, (2008b) se hace un estudio desde el área de la biomedicina en el que se compara la cobertura temática de PubMed, Scopus, Web of Science y Google Scholar. Además, también se evalúan los productos en función de su utilidad como herramientas de recuperación de información biomédica. En este artículo se concluye que PubMed es todavía la mejor herramienta para la recuperación óptima de información biomédica, así como que Scopus tiene una mayor cobertura de revistas que WoS, aunque limitada a artículos posteriores a 1995, y por último, que Google Scholar puede ayudar a encontrar recursos informativos más oscuros, pero tiene una peor, y menos frecuentemente actualizada información de citación que en las otras dos bases de datos de citas. Jacso (2005) también analiza la cobertura temática por de WoS, Scopus y Google Scholar, tanto por número de revistas indizadas, como por número de registros existentes en las bases de datos, y también se analiza el rango temporal que tiene en cuenta cada producto para contabilizar las citaciones. En sus conclusiones, arremete contra Google Scholar, criticando la imposibilidad de crear consultas estructuradas mediante lógica Booleana, y, sobre todo, sus limitaciones a la hora de procesar correctamente los trabajos científicos, cuando normalmente éstos están mucho más estructurados y contienen muchos más metadatos que los miles de millones de páginas web que se indizan en su buscador general. En Norris *et al.* (2007), el estudio se centra en las alternativas que hay a WoS desde el punto de vista de las Ciencias Sociales, y se estudia la cobertura que Scopus, CSA Illumina y Google Scholar hacen de las mismas. Concluye que las mejores opciones entre las cuatro son Scopus y Web of Science, y, debido a que Scopus es la base de datos que ofrece una mayor cobertura, propone ésta como alternativa a WoS.

Asimismo, también hay estudios que comparan los indicadores bibliométricos calculados a partir de datos de Google Scholar, ya sea a nivel de artículo (con las citas contabilizadas en GS), a nivel de investigador individual (como en GS Citations), o a nivel de revista (como en GS Metrics), con los indicadores que se pueden calcular con

datos de otros servicios, principalmente con WoS y Scopus. En Delgado López-Cózar *et al.* (2013) se hace un estudio de la correlación entre el factor de impacto de los JCR, el SJR de Scopus, y el índice h a 5 años de Google Scholar Metrics, donde se concluye que a pesar de las diferencias que tiene este último respecto de los otros dos, y de otras diversas limitaciones (falta de transparencia en su elaboración y presentación, falta de control contra la manipulación) los resultados que se obtienen en sus indicadores son muy similares a los de las fuentes tradicionales. Por último, también se pueden encontrar estudios comparativos por áreas de conocimiento, como el de Franceschet (2010) en el campo de las Ciencias de la Computación, donde se concluye que los rankings basados en indicadores de citación no cambian significativamente cuando se computan con los datos de una y otra fuente.

Sobre las técnicas de solapamiento entre fuentes secundarias de información científica (bases de datos científicas), merecen ser mencionados varios trabajos. El de Bearman *et al.* (1977) es el más extenso ya que abarca el solapamiento de catorce bases de datos, entre ellas Chemical Abstracts y Biological Abstracts. Fue de las primeras investigadoras en utilizar técnicas estadísticas como el escalado multidimensional (MDS) y el análisis factorial, aplicadas al solapamiento de datos. Por su parte, Gluck (1990), realiza una revisión sobre los estudios relacionados con el solapamiento, y además introduce una alternativa al TO (traditional overlap), el *Syndetic freedom* (SF). En Hood *et al.* (2003) se analiza el solapamiento entre documentos del campo de la teoría de conjuntos difusos, encontrados en todas las bases de datos disponibles del sistema de información DIALOG. Se concluye que el 63% del total de documentos solo se podían encontrar en una de las bases de datos. En el ámbito español podemos encontrar el trabajo de Escalona Fernández *et al.* (2010), donde además de hacer también una buena revisión bibliográfica de la literatura sobre el solapamiento de bases de datos científicas, realizan un estudio del solapamiento de las revistas de ingeniería química con los datos de la Web of Science y Scopus.

2. Objetivos

El objetivo principal de este trabajo es diseñar e implementar una base de datos que sea capaz de identificar unívocamente a las principales revistas científicas editadas en el mundo, a partir del principal sistema de información de publicaciones periódicas (Ulrich's) y de las principales bases de datos e índices de evaluación de revistas científicas (Web of Science-JCR, Scopus-SJR), integrando asimismo sus indicadores bibliométricos a fin de disponer de información puntual y actualizada. Concretamente, se pretende comenzar esta Master List integrando las siguientes bases de datos:

- Master Journal List de la Web of Science (*Science Citation Index Expanded*, *Social Sciences Citation Index*, y *Arts & Humanities Citation Index*) y *Journal Citation Reports* (SCI y SSCI).
- *Scopus Title List*, junto con el *SCImago Journal Rank* (en el que se ofrecen datos complementarios a la lista de fuentes).
- *Ulrich's International Periodicals Directory*. En este caso, sólo las revistas consideradas como académicas. El directorio Ulrich's es considerado como la fuente de información sobre publicaciones periódicos más exhaustiva del mundo, y ha sido utilizada en numerosas ocasiones en estudios de cobertura de bases de datos de información científica.

Por tanto, el centro de gravedad de este trabajo consistirá en describir todo el proceso de descarga, depuración y normalización que implica la creación de esta base de datos. Se describirán los problemas técnicos derivados de mismo y se obtendrán datos sobre el control bibliográfico ejercido por cada uno de los sistemas de información ejercidos.

Asimismo, se plantean como objetivos secundarios:

- Analizar la cobertura de las bases de datos: lingüística, geográfica, editorial (organismos editores)
- Determinar el solapamiento entre las tres bases de datos.
- Describir las características de las clasificaciones temáticas que utilizan estos productos.

3. Material y métodos

La creación de esta base de datos se puede dividir en tres fases:

1. Descarga de los datos brutos desde las respectivas páginas web de cada producto.
2. Procesamiento individual de cada producto, y conversión a base de datos relacional.
3. Integración de los productos en una única base de datos.

Seguidamente se detalla el proceso que se ha seguido para cada base de datos.

3.1. Ulrich's International Periodicals Directory

Descarga

Para la descarga de la base de datos de Ulrich's desde su interfaz web (<http://ulrichsweb.serialssolutions.com/>), se diseñó un programa que aprovechaba las

funcionalidades que ofrece esta base de datos para descargar registros. Ulrich's permite seleccionar hasta 100 registros de una consulta, y descargarlos en un formato compatible con hojas de cálculo. En el momento de la descarga, que se realizó en mayo de 2013, Ulrich's contenía 616.409 registros, que se podían visualizar realizando una consulta que contuviera solamente el operador de truncamiento (*) en su buscador.

En un principio, como lo que se buscaba era solamente las revistas científicas, se pensó que para ahorrar esfuerzo, la consulta se filtraría para descargar solo este tipo de revistas, pero esta opción se desechó al comprobar que Ulrich's cometía errores de indización que dejarían fuera a revistas que nos interesaban, como se puede ver en la figura 1.

En este ejemplo se puede ver que si en nuestra consulta se hubiera filtrado por Tipo de Serie = Revista, o por Tipo de contenido = Académico, se habrían perdido los registros de algunas revistas (sirva de ejemplo la versión electrónica de la *Revista Española de Cardiología*). Por tanto, como en principio no era posible saber con qué frecuencia Ulrich's cometía este tipo de errores, se decidió descargar la base de datos al completo, y posteriormente crear unos filtros personalizados para quedarnos solamente con las revistas científicas.

Figura 1. Ejemplo de errores de indización en Ulrich

| | Título | Editorial | ISSN | País | Estad. | Tipo de serie | Idioma | Tipo de contenido | Formato |
|--|---|----------------|-----------|--------|--------|---------------|---------|------------------------------|----------|
| | Revista Espanola de Cardiologia (Print) | Elsevier Doyma | 0300-8932 | España | Activa | Revista | Español | Académico / de investigación | Impreso |
| | Revista Espanola de Cardiologia. Suplemento | Elsevier Doyma | 1131-3587 | España | Activa | Revista | Español | Académico / de investigación | Impreso |
| | Revista Espanola de Cardiologia (Online) | Elsevier Doyma | 1579-2242 | España | Activa | Vacío !! | Español | Vacío !! | En línea |
| | Revista Espanola de Cardiologia. Suplemento | Elsevier Doyma | 1579-2250 | España | Activa | Revista | Español | Académico / de investigación | Impreso |
| | Revista Espanola de Cardiologia (Online) | Elsevier Doyma | 1885-5857 | España | Activa | Vacío !! | Inglés | Vacío !! | En línea |
| | Revista Espanola de Cardiologia. Suplemento | Elsevier Doyma | | España | Activa | Revista | Español | Académico / de investigación | En línea |

Si filtramos por Tipo de serie = Revista, o por Tipo de contenido = Académico / Inv., perderemos estos registros.

El programa que se diseñó para la descarga de los datos funcionaba como un *usuario automática*, descargando los 100 registros que se pueden visualizar y seleccionar por página, y navegando a la siguiente página. Esta operación se realizó 6.165 veces, produciendo el mismo número de archivos de hoja de cálculo.

Procesamiento

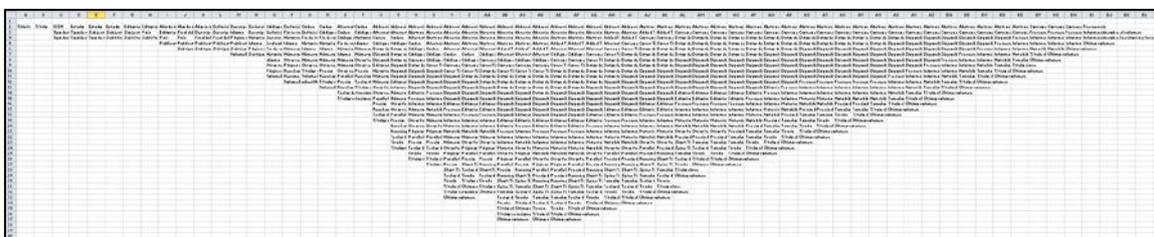
Como es normal, para su procesamiento, primero había que unir en uno todos los archivos que se habían descargado. Esta tarea resultó dar bastantes quebraderos de cabeza, porque sucedió algo que no había sido previsto.

La base de datos de Ulrich's contiene un gran número de campos, 71 para ser exactos, pero no todos ellos se utilizan en todos los registros. Por esta razón, cuando en cualquiera de las descargas que se habían realizado (recordemos, 6.165 archivos

con 100 registros por archivo) daba la casualidad de que el campo x no se había utilizado en ninguno de esos 100 registros, dicho campo se omitía en la descarga. La consecuencia de esto fueron discrepancias en las columnas que ocupaban los campos, entre los 6.165 archivos descargados.

Esto convirtió la tarea del procesamiento de estos datos en algo más complicado que simplemente unir cada tabla una debajo de otra. Para resolver este problema se diseñó un programa que analizaba la primera fila de todas las columnas de cada archivo, es decir, la cabecera, donde venía el título de los campos, y que, tras analizar el título del campo, copiaba cada columna y la pegaba en la columna adecuada de otra hoja de cálculo donde se habían establecido unas posiciones fijas para cada uno de los 71 campos de Ulrich's.

Figura 2. La fosa. Parte del proceso de procesamiento de los datos de Ulrich's, donde se determinaba qué campos podían aparecer en una misma columna de la hoja de cálculo



Una vez unidos todos los archivos en una sola hoja de cálculo de 616.409 registros, se procedió a la eliminación de duplicados, momento en el cual detectamos otro imprevisto. En la interfaz de consulta de Ulrich's se podían recuperar tanto publicaciones activas, es decir, que se siguen editando en la actualidad, como inactivas, porque o bien han muerto, o han sufrido alguna transformación (se han cambiado de nombre, o se han fusionado con otras publicaciones, o lo contrario, se han separado en varias publicaciones). Sin embargo, en el proceso de descarga, todas aquellas publicaciones inactivas que no lo fueran por haber sido descontinuadas, sino por haber sufrido alguno de los cambios descritos anteriormente, se remitían al registro más actual, es decir, al que actualmente consta como activo, y era este el que se descargaba. Esto no quiere decir que se perdiera la información sobre los títulos anteriores de las actuales revistas activas, porque esa información seguía almacenada en el campo *History* del registro activo. Sin embargo, sí quiere decir que en nuestra lista había un gran número de duplicados “exactos”, que se eliminaron fácilmente al quedarnos solo con una copia de cada *TitleId*, (el campo que utiliza Ulrich's como clave principal). Tras esto, quedaron 505.574 registros.

Una vez hecho esto, se debían detectar los duplicados “no exactos” es decir, aquellos que, aunque Ulrich's reconocía como registros diferentes, realmente son la misma revista. Un ejemplo de ellos lo tenemos en la figura 1, donde vemos que Ulrich's almacena por separado la versión impresa y la versión online de la *Revista Española de Cardiología*. Para hacer esto se procesó el campo *Título*, donde, además del título de la publicación, aparece también el formato de edición. Por ejemplo:

“Revista Española de Cardiología (Print)” y “Revista Española de Cardiología (Online)”. Se creó un nuevo campo de Título sin el añadido del formato, y se estableció que los duplicados serían aquellas revistas que tuvieran los mismos títulos, y, a la vez, que estuvieran publicadas por la misma editorial, para resolver el caso de posibles revistas homónimas. Ahora bien, antes de borrar los duplicados, primero se añadió alguna información complementaria proveniente de estos duplicados, al registro “principal”, como el ISSN, el medio de publicación (impreso/online), el tipo de contenido (recordemos, a nosotros nos interesaban las revistas académicas) y el *TitleId* por si alguna vez hiciera falta recuperar más datos del registro duplicado. Tras este proceso, quedaron 380.010 registros.

A continuación, se prosiguió con la identificación de las revistas científicas. Para esto, en principio se cogieron todas las publicaciones que Ulrich’s consideraba revistas académicas (es decir, lo que habíamos considerado descargar en un principio), pero además, tras hacer una lista de todos los ISSN y títulos de revistas que aparecieran en las bases de datos de WoS y de Scopus (que ya teníamos procesadas), y de enfrentarlas con estos mismos campos en Ulrich’s, se recuperaron también todos aquellos registros que, aun no siendo considerados por Ulrich’s como revistas académicas, se encontraban en alguna de estas otras dos bases de datos. En este proceso, finalmente quedaron 68.999 registros.

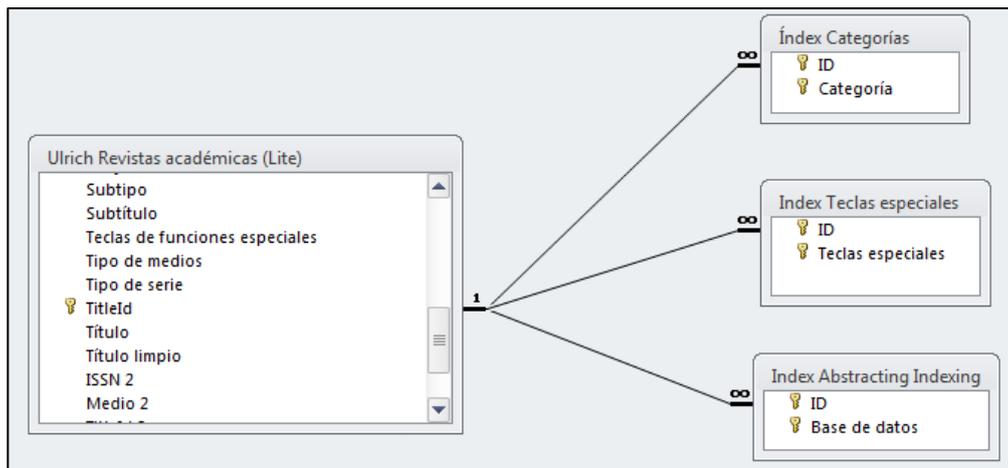
Por último, se crearon los índices para los campos:

- *Abstracting/Indexing*: donde se consigna una lista de las bases de datos donde está indizada cada revista.
- *Teclas de funciones especiales*: en este campo con título tan poco descriptivo encontramos un cajón de sastre donde se almacena información importante, como si la revista tiene revisión por pares, si es una publicación *Open Access*, o si está indizada en los JCR.
- *Subject Codes*: donde se guardan las categorías temáticas a las que pertenece cada revista.

Hay más campos de los que se podrían haber creado índices, pero para los propósitos de este trabajo se ha considerado que estos son suficientes. Sin embargo, siempre existe la posibilidad de hacerlo en el futuro, ya que todos los demás campos se han incluido en la tabla principal como campos de texto plano.

La estructura de relaciones resultante tras todo este procesamiento es la siguiente:

Figura 3. Mapa de relaciones de Ulrich's



3.2. WoS Master Journal Lists / Journal Citation Reports.

Descarga

Se desarrollaron una serie de aplicaciones ad-hoc con el objetivo de descargar los datos brutos de estas fuentes. Estas aplicaciones utilizaban la técnica del *parsing*, similar a la que utilizan los robots de los motores de búsqueda. Estos programas navegaban por las interfaces de búsqueda de estos productos y rastreaban su código HTML extrayendo la información relevante de cada revista. Estos datos eran posteriormente almacenados en un formato compatible con hojas de cálculo y sistemas integrados de gestión de bases de datos (en concreto, archivos de texto tabulados).

En el caso de la *Master Journal List* de Thomson Reuters (edición de octubre de 2013), ésta posee una interfaz de consulta convencional (accesible desde la siguiente dirección: <http://ip-science.thomsonreuters.com/mjl/>), y no ofrece ninguna funcionalidad de exportación de datos, por lo que en este caso el uso de un *parser* era la única opción para extraer los datos. De todos los productos de Thomson Reuters, se descargaron la lista de revistas de las bases de datos que integran la *Web of Science: Arts & Humanities Citation Index*, *Science Citation Index Expanded*, y *Social Sciences Citation Index*. Además, el listado de revistas de estos tres productos también se puede consultar a través de tres archivos PDF que se encuentran en la misma dirección. El listado de revistas es el mismo, pero en estos archivos aparece un dato interesante que no aparece en el buscador: el ISSN electrónico de las revistas, cuando lo tienen. Para no perder esta información, también se descargaron los archivos PDF, y se convirtieron a hojas de cálculo para su posterior procesamiento. La descarga se realizó el día 18 de octubre de 2013.

Por su lado, en los *Journal Citation Reports* (<http://webofknowledge.com/>), aunque la interfaz ofrece algunas funciones para la exportación de los datos, éstas no

Para representar las Masters Lists del SSCI, SCI, y A&HCI, solo hicieron falta una tabla para los datos bibliográficos de las revistas, y otra para sus categorías. Además, las listas de categorías del SSCI y del SCI son las mismas que las de sus respectivos JCR, así que se decidió no duplicarlas ya que no era necesario.

Como se puede apreciar en la figura 4, en el caso de los JCR, los datos bibliométricos, los de cuartiles y las categorías se han almacenado en tablas separadas de los datos bibliográficos. Este diseño responde a la necesidad de que sea fácil actualizar la base de datos cuando aparezcan nuevas ediciones de los JCR.

En la tabla de indicadores (*JCR_SSCI_INDICATORS* y *JCR_SCI_INDICATORS*) la clave principal está compuesta de tres campos, que son el título abreviado, el ISSN, y otro campo al que se ha llamado *JCR_EDITION*. Los dos primeros son necesarios para identificar unívocamente la revista, pues cada uno por separado no es capaz de hacerlo. En *JCR_EDITION* se consigna el año al que corresponden los indicadores que aparecen en dicho registro. Por tanto en esta tabla se podrán agregar los factores impacto que se vayan calculando para las sucesivas ediciones de los *Journal Citation Reports*. Para comprobar que este modelo funciona, se han agregado los datos de impacto de los JCR 2011, y 2012.

En el caso de los cuartiles y las categorías, también se han almacenado en tablas diferentes (*JCR_SSCI_CAT_J* y *JCR_SCI_CAT_J*) porque es la única manera adecuada de poder consignar un cuartil por cada categoría temática en la que se encuentra, y por cada año que la revista aparezca en los JCR. Por esta razón, en estas tablas, la clave principal está formada por cuatro campos: título e ISSN para identificar la revista, *JCR_EDITION* para identificar cada año, y *COD_CAT_** para identificar a cada categoría.

3.3. Scopus Title List / SCImago Journal Rank

Descarga

Tanto la lista de fuentes de Scopus (<http://www.elsevier.com/online-tools/scopus/content-overview>), como el ranking de revistas de SCImago (<http://www.scimagojr.com/journalrank.php>) se descargaron sin mucha dificultad, ya que desde sus respectivas páginas web están disponibles para su descarga en formato hoja de cálculo. Las descargas se realizaron en diciembre de 2013.

La única limitación que se encontró fue que si se descargaba el ranking completo de SCImago, en el archivo resultante no aparecían las áreas y categorías a las que están asignadas las revistas. Considerando ésta una información valiosa, para poder guardar esta información, se descargaron las 307 categorías una a una mediante una aplicación diseñada al efecto. También se extrajo la lista de áreas temáticas, junto con la lista completa de categorías asociadas a cada área, para poder construir posteriormente la clasificación completa en dos niveles. Al hacer esto se vio

que nos encontrábamos ante una clasificación jerárquica simple, en la que cada categoría pertenece únicamente a un área temática. Solo se han encontrado tres categorías que no obedecen a esta norma, ya que no están asignadas a ningún área. Son: *Social Work*, *Nanoscience and Nanotechnology*, y *Medical Assisting and Transcription*.

También cabe reseñar que, aunque tanto en la lista de fuentes de Scopus, como en el SJR, no solo hay revistas científicas, sino también otras tipologías documentales como actas de congresos, en la base de datos se han incluido todos los registros, aunque luego solo se tendrán en cuenta las revistas para el estudio.

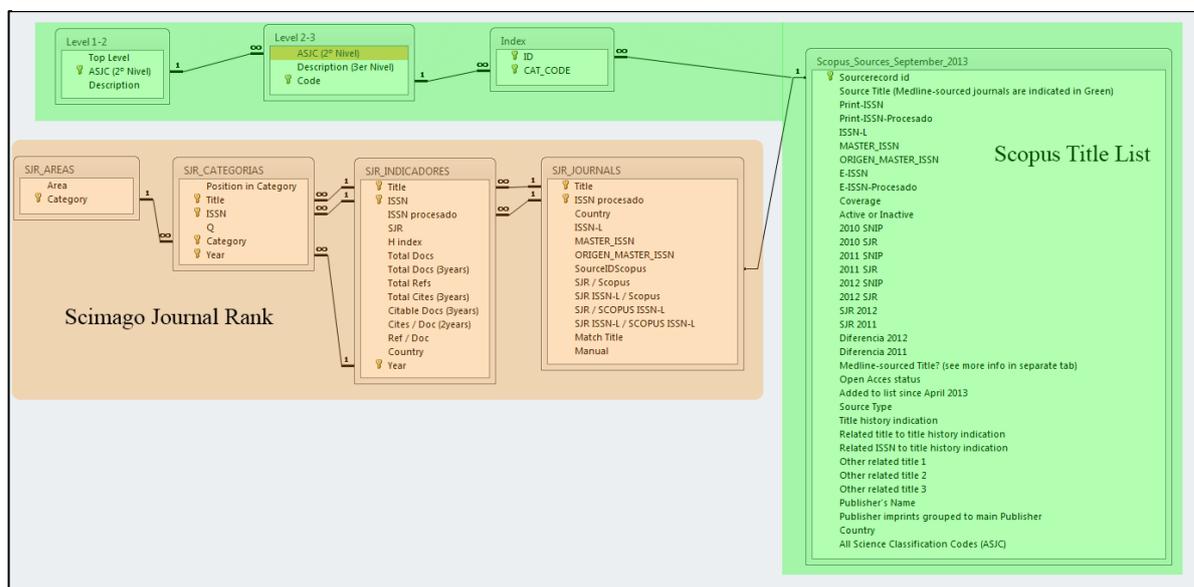
Procesamiento

La unificación entre los registros de la lista de fuentes de Scopus, y el ranking de SCImago se hizo mediante los campos del ISSN, el ISSN-L, y el título, en este orden de precedencia. Se consiguió un emparejamiento SJR/Scopus casi perfecto, en el que el 99,99% de los elementos que se encontraban en el SJR, se encontraron también en Scopus, lo que es normal si pensamos que el primero está basado en datos del segundo. Solamente se quedaron descolgados 11 registros (de 20544) del SJR que no fueron encontrados en Scopus. Tampoco fue posible encontrar coincidencias de manera manual para estos 11 registros.

Una vez hecho el emparejamiento, solo se añadieron a cada producto sus respectivas clasificaciones temáticas en formato relacional.

En este caso no se creó una tabla maestra como se hizo con los productos de WoS, porque ya ha quedado establecido que todas las revistas del SJR también se encuentran en la lista de fuentes de Scopus, excepto por el pequeño remanente del que se ha hablado.

Figura 5. Mapa de relaciones Scopus / SJR



Como se puede apreciar en la figura 4, aquí se hay utilizado una práctica similar a la que se utilizó en los JCR para poder actualizar los datos bibliométricos conforme aparezcan nuevas ediciones de los indicadores.

Tenemos una tabla de indicadores, en la que en la clave principal se encuentran los campos necesarios para identificar a la revista, y otro campo para identificar el año al que se refieren los indicadores (*Year*).

A su vez, también tenemos otra tabla en la que se consigna el cuartil en el que se encuentra cada revista en cada categoría, y en cada año.

3.4. Integración de las tres bases de datos.

Sin duda, el principal reto de este trabajo, una vez descargados y procesados los productos individualmente, es unir los datos de una revista en un producto, con los datos de la misma revista en los demás productos. Esta labor, a la que nos referiremos como *matching* o emparejamiento, es problemática debido a la falta de normalización bibliográfica.

Como sabemos, todas las revistas, como publicaciones periódicas que son, deben poseer un código identificativo llamado ISSN, que son las siglas de *International Standard Serial Number*. Sin embargo, este identificador no es asignado a la revista como entidad abstracta, sino que se asigna a cada medio de publicación en el que se publique la obra. Así, una publicación como la *Revista Española de Cardiología*, que es publicada tanto en formato impreso como online, tiene un ISSN impreso (0300-8932), y un ISSN electrónico o e-ISSN (1579-2242). Además, como también es publicada en inglés en formato electrónico, posee otro más (1885-5857). Así, en una base de datos podríamos encontrar solamente el ISSN impreso, y en otra solamente el ISSN electrónico, y en este caso el *matching* a través de este campo sería imposible.

Los títulos de las revistas tampoco se escapan de los problemas de normalización. La casuística es bastante extensa, pero sirvan estas tres muestras a modo de ejemplo:

- El *British Medical Journal* podría aparecer en una base de datos con ese título, pero en otra puede aparecer como *BMJ*.
- *The Spanish Journal of Psychology* podría aparecer en una base de datos con ese nombre, pero en otras como *Spanish Journal of Psychology*, o como *Spanish Journal of Psychology (The)*.
- En el caso de las revistas españolas, algunos caracteres de nuestro idioma también suelen suponer una barrera a la normalización. La *Revista Española de Cardiología*, tomándola de nuevo como ejemplo, puede aparecer en las bases

de datos internacionales como *Revista Espanola de Cardiologia*, es decir sin acentos y sin ñ.

Al comenzar a realizar este estudio, solo contamos con estos dos campos, ISSN y título, para intentar hacer un matching automático mediante equiparación exacta. También se estudió la posibilidad de aplicar alguna técnica de equiparación parcial, para hacer uniones incluso si los títulos no eran exactamente iguales, aunque sí parecidos. Estas técnicas eran:

- Calcular la *distancia de Levenshtein*, que es el número mínimo de transformaciones requeridas para convertir una cadena de texto x en otra cadena y . Las transformaciones posibles son inserción, eliminación o sustitución de caracteres. Así, cadenas de texto (en este caso títulos de revistas) diferentes pero similares tendrían una distancia pequeña entre ellos, y esto facilitaría determinar si realmente son la misma revista o no. Esta opción se desestimó por requerir que cada título de revista de una base de datos tuviera que ser comparado con todos los títulos de revistas en las demás bases de datos, creando así matrices de miles o decenas de miles de filas y columnas, lo cual sobrepasaba las capacidades de procesamiento de un PC normal, y de los programas de procesamiento de datos convencionales.
- Crear un pequeño y simple sistema de recuperación de información, en el que los documentos que se almacenaran en el fichero inverso fueran los títulos de las revistas de una base de datos (las palabras que forman el título, aunque también se podrían utilizar *n-gramas*), y que automáticamente fuera consultado con los títulos de revistas de otra base de datos. Esta consulta se procesaría con un *tf-idf* simple, respondiendo con una lista de títulos de revista similares al que se hubiera buscado. Esta opción tampoco se materializó por no conocer ningún sistema de estas características que pudiera implementarse de manera más o menos sencilla para poder llevar a cabo esta tarea.

Sin embargo, sí se encontró una opción que prometía ayudar a resolver los problemas de normalización entre los ISSN de las revistas: el ISSN-L, también llamado ISSN de enlace.

El ISSN-L fue introducido en la revisión de la norma del ISSN que se publicó en agosto de 2008 (ISO 3297-2007), y fue creado “para permitir la utilización del ISSN cuando se quiere encontrar o enlazar a un recurso continuo independientemente de su soporte”. El ISSN-L no es un nuevo código asignado a cada una de las publicaciones periódicas existentes, sino que uno de los diferentes ISSN asignados a una misma revista cumpliría también la función de ISSN-L. En concreto, el ISSN-L sería el primer ISSN atribuido en el Registro del ISSN a un recurso continuo, independientemente del medio al que correspondiera ese primer ISSN. El ISSN-L se asigna a todas las publicaciones periódicas, incluso si dicha revista existe solamente

en un solo soporte, y solamente se asigna un ISSN-L para cada recurso continuo, sin importar el número de versiones en diferentes soportes que existan de dicha publicación.

Algunos ejemplos:

1. Recurso continuo en un solo soporte físico:

ISSN-L 1748-7188

Recurso en línea: *Algorithms for molecular biology* = ISSN 1748-7188

2. Recursos continuos en dos soportes físicos:

ISSN-L 0264-2875

Versión impresa: *Dance research* = ISSN 0264-2875

Versión en línea: *Dance research* (Online) = 1750-0095

3. Recursos continuos en tres soportes físicos:

ISSN-L 1188-1534

Versión impresa: *Plant varieties Journal* = 1188-1534

Versión en línea: *Plant Varieties journal* (Online) = 1911-1479

Versión en CD-ROM: *Plant varieties journal* (CD-ROM) = 1911-1460

En teoría, el ISSN-L debería resolver los problemas de normalización de ISSN entre las diferentes bases de datos que queremos integrar. Afortunadamente, aunque en las bases de datos que se han descargado no aparece este ISSN-L, la agencia internacional del ISSN proporciona, de forma gratuita, una **tabla de correspondencias ISSN/ISSN-L** a todo aquel que la solicite desde su página web. La versión que obtuve al pedir esta lista fue la del día 19 de septiembre de 2013.

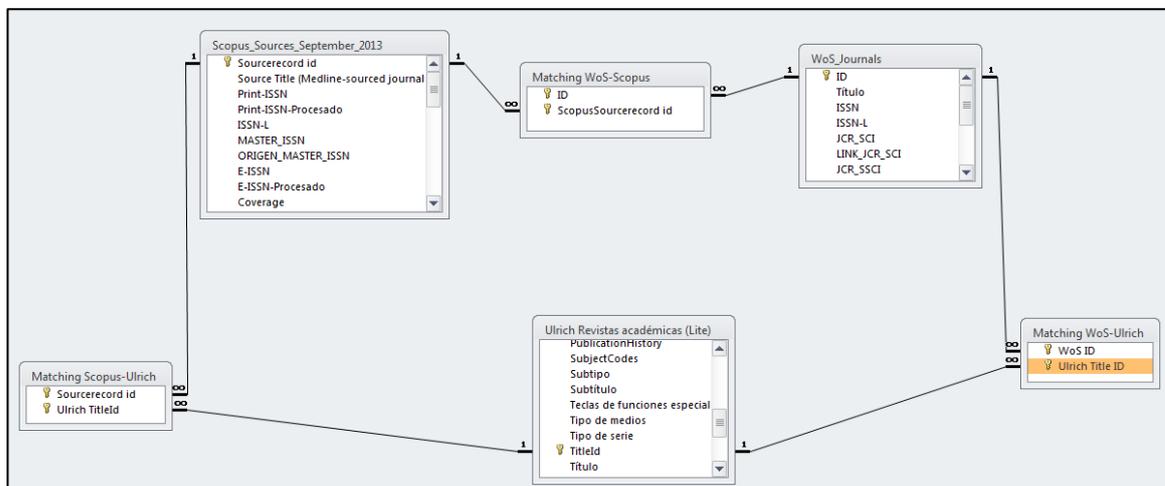
Teniendo esta tabla de correspondencias, fue sencillo crear en cada una de nuestras bases de datos un campo llamado ISSN-L, y, mediante consultas de actualización en las que relacionábamos los campos de ISSN que proporcionan las propias bases de datos, y la tabla de correspondencias ISSN/ISSN-L, asignar los ISSN-L a las revistas de WoS, Scopus, y Ulrich's.

Una vez hecho esto, se procedió a realizar el *matching* entre las tres bases de datos, primero de dos en dos, y finalmente las tres juntas ($WoS \cap Scopus$, $WoS \cap Ulrich's$, $Scopus \cap Ulrich's$, y $WoS \cap Scopus \cap Ulrich$). Para realizar el *matching* se combinaron los campos de ISSN (el original de la base de datos), el ISSN-L, y, en última instancia, si los dos métodos anteriores fallaban, el campo de título.

En el caso del *matching* entre la lista de fuentes de Scopus y Ulrich's, también se utilizaron en el proceso los ISSN que aparecen en el campo *History* de Ulrich's, ya que entre las revistas de Scopus se encuentran un gran número de revistas inactivas que no se podían emparejar con la lista principal de Ulrich's, pues en esta lista, como se ha dicho anteriormente, solo aparecen las versiones más recientes de cada revista.

El mapa de relaciones, tras este proceso, es el siguiente. En él están visibles solo las tablas principales de cada producto, pero las relaciones que se mostraban anteriormente por separado, continúan siendo las mismas.

Figura 6. Mapa de relaciones entre Ulrich's, WoS y Scopus



4. Resultados

4.1. Resultados generales. Activas/Inactivas, y control del ISSN y de títulos

El total de registros que se han descargado de cada producto, así como el total de revistas que han quedado tras el procesamiento, se puede ver en la tabla 1. También se pueden observar los subconjuntos de revistas activas (es decir, revistas que el producto sigue indizando en la actualidad), y las inactivas (revistas que han muerto, se han transformado, o que el producto ha decidido dejar de cubrir). Los porcentajes están calculados respecto del total de revistas académicas (3ª columna).

Tabla 1. Número de registros bibliográficos de revistas descargadas por las bases de datos, distribuidas por tipo de publicación y grado de actividad

| Base de datos | Registros descargados | Revistas académicas | Situación actual de las revistas | | | | | |
|---------------------|-----------------------|---------------------|----------------------------------|--------------|--------------|--------------|-------------|-------------|
| | | | Activas | % | Inactivas | % | Desconocido | % |
| Ulrich's | 616409 | 68999 | 61147 | 88,6% | 5085 | 7,4% | 2767 | 4,0% |
| WoS/JCR | 12779 | 12779 | 12779* | - | - | - | - | - |
| WoS | 12490 | 12490 | 12490* | - | - | - | - | - |
| JCR | 11145 | 11145 | 10925* | 98,0% | 220 | 2,0% | 0 | 0,0% |
| Scopus / SJR | 33657 | 31270 | 20698 | 66,2% | 10572 | 33,8% | 0 | 0,0% |
| Scopus | 33646 | 31270 | 20698 | 66,2% | 10572 | 33,8% | 0 | 0,0% |
| SJR | 20544 | 19841 | 19751 | 99,5% | 79 | 0,4% | 11 | 0,1% |

Como era de esperar, el directorio Ulrich's es el producto con un mayor número de revistas (61147). El resto de las revistas académicas que aparecen en el directorio están inactivas (5085) o se desconoce su estado (2767), lo que ocurre cuando no se especifica este dato en el registro de la revista.

En el caso de las revistas de la Master List de WoS, y las revistas del JCR, no se ha podido averiguar el número de revistas inactivas (es decir, las que se cubrían en el pasado, y ya no), ya que Thomson Reuters no hace pública esa información. Se han tomado como revistas activas aquellas que aparecen en la edición de 2012 de los JCR o en la versión más reciente de las Master Journal Lists, que data de octubre de 2013. Al contrario que WoS, Scopus es mucho más transparente a la hora de ofrecer esta información, la cual aparece en el archivo de la Scopus Title List, libremente accesible en su web. Del total de revistas que aparecen en este fichero, dos tercios (20698) son revistas activas. En el caso de las revistas que aparecen en los listados del SJR, el 99,5% son activas (19751).

En la tabla 2 se muestran algunos datos acerca del uso que hace cada producto del ISSN. Aquí los porcentajes también se refieren al total de revistas académicas de cada producto. Las tres bases de datos incluyen alguna versión del ISSN en la mayoría de sus registros, aunque WoS/JCR es la que hace un uso más extenso, con un 99,98% de revistas que presentan ISSN (solo hay dos revistas que no lo tengan). Además, en las Master Journal Lists también se muestran los ISSN electrónicos (no así en los JCR). En Scopus, la mayoría de las revistas (97,7%) también incluyen la información sobre el ISSN impreso, y aproximadamente el 24% también el ISSN electrónico. Aun así, hay 718 revistas que no presentan ningún tipo de ISSN (2,3%). Por último, el directorio Ulrich's posee ISSN impreso en el 83% del total de revistas, ISSN electrónico para el 43%, y un 5,5% de revistas en las que no contiene ningún ISSN. SJR tampoco ofrece ISSN electrónicos, aunque podrían recuperarse del listado de fuentes de Scopus.

Tabla 2. Control del ISSN

| Base de datos | Revistas académicas | Control del ISSN | | | | | | | |
|-------------------------|---------------------|------------------|---------------|--------------|--------------|-------------|--------------|-------------|-------------|
| | | ISSN impreso | | ISSN elect. | | sin ISSN | | sin ISSN-L | |
| | | | % | | % | | % | | % |
| Ulrich's | 68999 | 57387 | 83,17% | 29835 | 43,2% | 3817 | 5,53% | 5880 | 8,5% |
| Master Lists/JCR | 12779 | 12777 | 99,98% | 10035 | 78,5% | 2 | 0,02% | 141 | 1,1% |
| Master Lists | 12490 | 12488 | 99,98% | 10035 | 80,3% | 2 | 0,02% | 134 | 1,1% |
| JCR | 11145 | 11143 | 99,98% | - | - | 2 | 0,02% | 117 | 1,0% |
| Scopus / SJR | 31270 | 30552 | 97,70% | 7439 | 23,8% | 718 | 2,30% | 1240 | 4,1% |
| Scopus | 31270 | 30552 | 97,70% | 7439 | 23,8% | 718 | 2,30% | 1240 | 4,1% |
| SJR | 19841 | 19622 | 98,90% | - | - | 216 | 1,09% | 537 | 2,7% |

También cabe destacar que en el proceso de asignación de los ISSN-L a cada base de datos mediante la tabla de correspondencias ISSN / ISSN-L, hay algunos casos en los que no se recuperaron resultados. En Ulrich's hubo 5880 ISSN que no se encontraron en la tabla de correspondencias (8,5%), mientras que del conglomerado WoS/JCR no se encontraron 141 (1,1%), y de Scopus, 1240 (4,1%). Estos hechos reflejan que ni siquiera el Centro Internacional del ISSN tiene un control perfecto de las publicaciones periódicas con ISSN, ya que en la tabla de correspondencias no están todos los ISSN, a pesar de ser una tabla con más de 1,7 millones de registros.

El control de los diferentes títulos que puede recibir una misma revista en cada base de datos ha resultado más complicado, ya que la casuística de las diferencias que pueden existir es muy grande. Para estudiar las similitudes y diferencias entre títulos se ha optado con un método cuantitativo como es el cálculo de la distancia de Levenshtein entre los los títulos que da cada base de datos a una misma revista. Una vez integradas las tres bases de datos, se ha calculado esta distancia, que mide el número de caracteres que diferencia a una cadena de texto, de otra. Con la distancia de Levenshtein se ha calculado un indicador de similitud normalizado (ISN) con la siguiente fórmula:

Siendo L = distancia de Levenshtein; TA = tamaño de la cadena A (número de caracteres); TB = tamaño de la cadena B; y MAX(TA;TB) = el tamaño de la cadena más grande de las dos:

$$ISN = 1 - (L / MAX(TA;TB))$$

Este índice puede tomar valor de 0 a 1 (0 cuando las cadenas son totalmente diferentes, 1 cuando son exactamente iguales).

Por ejemplo, si comparamos las cadenas: "Journal of Foot & Ankle Surgery" y "Journal of Foot and Ankle Surgery" vemos que son iguales excepto porque en una se representa la conjunción "y" como "&" y en la otra como "and". La distancia de Levenshtein entre estas dos cadenas es de 3, porque hay que hacer una sustitución

("&" por "a"), y dos adiciones de caracteres ("n" y "d") para convertir una cadena en la otra. El tamaño en número de caracteres de la primera cadena es 31, y el de la segunda 33, con lo cual el ISN sería:

$$\text{ISN} = 1 - (3 / \text{MAX}(31;33)) = 1 - (3 / 33) = 0,91$$

Por tanto, las dos cadenas coinciden en un 91%. En el Anexo 1 se exponen algunas de las casuísticas que se han encontrado sobre títulos diferentes.

En la tabla 3 se puede ver un resumen de las similitudes entre los títulos que las bases de datos dan a las revistas. En la primera columna (Distancia 0) se muestra el número de títulos que eran iguales en las dos bases de datos, y en la segunda, todos aquellos que tenían una distancia superior a 0, es decir, eran diferentes. El porcentaje es respecto del total de revistas activas. Todas las demás columnas tratan únicamente con aquellos títulos que eran diferentes entre sí.

Tabla 3. Resumen de las diferencias entre títulos

| | Dist. 0 | Dist. > 0 | % | Promedio | Moda | Límite inferior de los cuartiles | | |
|--------------------------|---------|-----------|--------|----------|------|----------------------------------|------|------|
| | | | | | | Q1 | Q2 | Q3 |
| WoS - Scopus | 9432 | 2513 | 26,64% | 0,754 | 0,9 | 0,95 | 0,91 | 0,54 |
| WoS - Ulrich's | 9299 | 2966 | 31,90% | 0,683 | 0,9 | 0,93 | 0,82 | 0,46 |
| Scopus - Ulrich's | 14401 | 4983 | 34,60% | 0,649 | 0,9 | 0,91 | 0,76 | 0,39 |

El promedio de la similitud de los títulos va desde el 65% que se consigue entre Scopus y Ulrich's, al 75% que se consigue entre WoS y Scopus. La moda, redondeada siempre al primer decimal, es 0,9, lo que quiere decir que la mayoría de las revistas tienen una similitud entre 0,85 y 0,95. Las columnas restantes muestran los límites inferiores de los cuartiles de las similitudes. Estos valores indican que la mayoría de las similitudes son altas, es decir, en la mayoría de los casos, la distancia de Levenshtein es muy baja entre las dos cadenas. En el caso de WoS y Scopus, casi el 50% de los títulos que no eran exactamente iguales, tienen un 90% o más de similitud entre ellos (mediana = 0,91).

4.2. Análisis de las clasificaciones temáticas

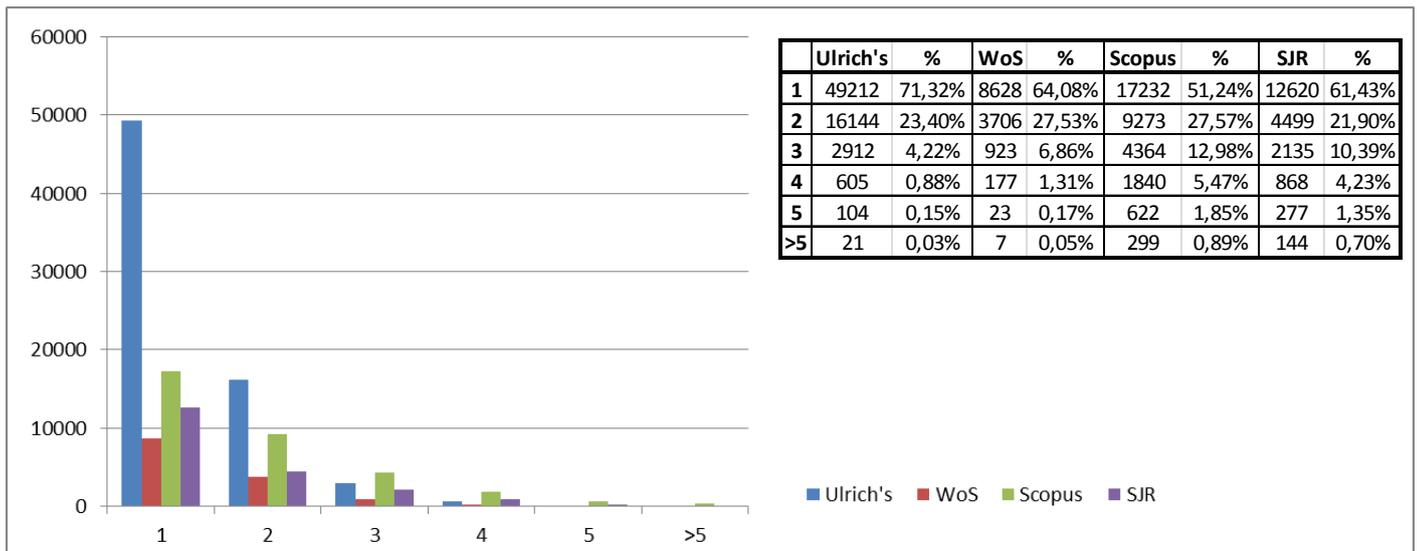
Las clasificaciones temáticas en estos tres productos son muy diversas. En la tabla 4 se puede ver un resumen de sus características.

Tabla 4. Características de las clasificaciones temáticas

| Base de datos | Niveles de clasificación | | Descripción |
|---------------|--------------------------|------------------|--|
| Ulrich's | 1 | Nivel único | 513 categorías, en algunos casos utilizando estructuras compuestas similares a los encabezamientos de materia. |
| WoS / JCR | 2 | Nivel General | 3 áreas (Science, Social Sciences, y Arts & Humanities). |
| | | Nivel Específico | 261 categorías entre las 3 áreas. |
| Scopus | 3 | Nivel General | 5 agrupaciones generales (General, Life Sciences, Physical Sciences, Health Sciences, Social Sciences). |
| | | Nivel Intermedio | 27 áreas. |
| | | Nivel Específico | 307 categorías. |
| SJR | 2 | Nivel General | Son los mismos que en el nivel intermedio y específico de Scopus, excepto contadas excepciones. |
| | | Nivel Específico | |

En las tres bases de datos, las revistas pueden ser asignadas a más de una categoría (entendiendo por categorías, el nivel más específico que posee el sistema de clasificación de cada base de datos). Se han querido estudiar las frecuencias con las que las revistas son asignadas a una o más categorías. En la gráfica 1 y su tabla asociada se puede observar que Ulrich's es la que menos categorías asigna en general a las revistas, mientras que Scopus y SJR tienden a asignar más.

Gráfica 1. N° de categorías asignadas por revista, en Ulrich's, WoS, Scopus, y SJR



4.3. Estudio del solapamiento

Los resultados del *matching* entre bases de datos, que aparecen en la tabla 5, se centran en las revistas activas de cada producto. En esta tabla se muestran los resultados del *matching*, detallando los emparejamientos que se han conseguido con cada uno de los campos que intervenían en el *matching*. La columna “ISSN” se refiere a los emparejamientos realizados solamente con los datos de ISSN que proporcionan los productos. La columna “Título” e “ISSN-L” funcionan de igual manera. En la columna “ISSN U Tít.” se hace la unión de los dos anteriores, para saber cuántos emparejamientos se habrían hecho, de no haber utilizado también el ISSN-L.

Como se puede observar, el ISSN-L por sí solo consigue peores resultados que la unión entre el ISSN y el título proporcionados por la base de datos (ISSN U Tít.), y solo proporciona unas decenas más de emparejamientos que no se podrían haber conseguido de otra manera (observando la diferencia entre el total de *matchings*, y los de la columna “ISSN U Tít.”). Esto significa que el ISSN-L no es una herramienta 100% efectiva para realizar esta tarea, aunque pueda utilizarse de forma complementaria a otras técnicas.

El resultado total aparece en la última columna, y es el mismo resultado que en la columna “Total (aut.)”, excepto en el caso de la intersección entre WoS y Scopus, donde además del emparejamiento automático, también se llevó a cabo un repaso manual en el que se realizaron 117 emparejamientos que no se habían encontrado automáticamente.

Tabla 5. Resultados del *matching* entre bases de datos, diferenciando por campos utilizados en el *matching*.

| Matching por campos | ISSN | Título | ISSN U Tít. | ISSN-L | Total (aut.) | Manual | Total |
|---------------------|-------|--------|-------------|--------|--------------|--------|-------|
| WoS ∩ Scopus | 11813 | 9432 | 11930 | 11751 | 11945 | 117 | 12062 |
| WoS ∩ Ulrich's | 12113 | 9299 | 12228 | 12076 | 12265 | - | 12265 |
| Scopus ∩ Ulrich's | 19081 | 14401 | 19351 | 18922 | 19384 | - | 19384 |

Para el estudio del solapamiento se utiliza el TO (*traditional overlap*), propuesto por Gluck (1990), y el RO (*relative overlap*) propuesto por Bearman *et al.* (1977).

Tomando dos bases de datos A y B:

$$TO = 100 \times \{ (A \cap B) / (A \cup B) \}$$

$$RO \% \text{ overlap en A} = 100 \times \{ (A \cap B) / A \}$$

$$RO \% \text{ overlap en B} = 100 \times \{ (A \cap B) / B \}$$

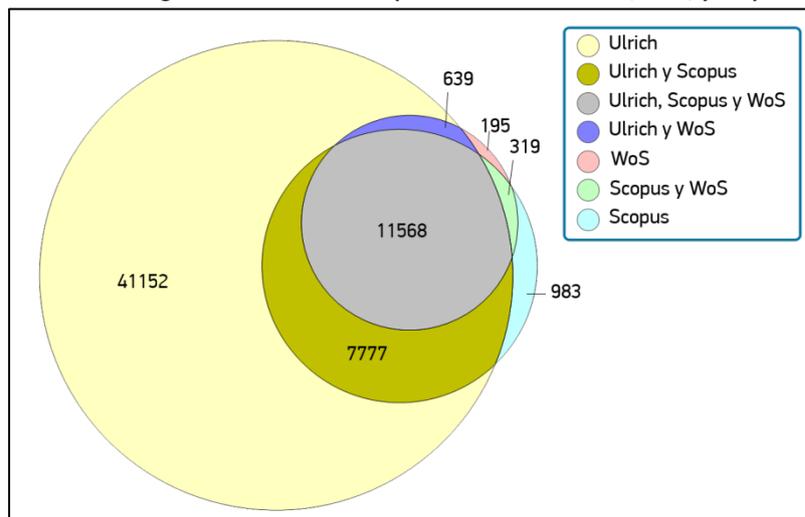
Tabla 6. Tradicional overlap (TO) y relative overlap (RO) entre Ulrich's, WoS y Scopus

| A | B | TO | RO (% en A) | RO (% en B) |
|--------|--------|--------|-------------|-------------|
| WoS | Ulrich | 19,88% | 95,98% | 20,06% |
| Scopus | Ulrich | 31,02% | 93,65% | 31,70% |
| WoS | Scopus | 56,00% | 94,39% | 58,28% |

El *traditional overlap* se puede interpretar de la siguiente manera, tomando como ejemplo WoS y Ulrich: teniendo en cuenta el total de revistas que se pueden encontrar tanto en WoS como en Ulrich's ($WoS \cup Ulrich$), hay un 19,88% de revistas que se encuentran en las dos bases de datos ($WoS \cap Ulrich$).

Por su lado, el *relative overlap*, tomando como ejemplo WoS (A), y Scopus (B), vemos que el 94,39% de WoS, está incluido en Scopus. Si intercambiamos el orden, el resultado es que el 58,28% de las revistas de Scopus se encuentran incluidas en WoS.

Gráfica 2. Diagrama de Venn del solapamiento entre Ulrich's, WoS, y Scopus



La intersección entre las tres bases de datos ($Ulrich's \cap WoS \cap Scopus$), tal y como se puede ver en la gráfica 1, es de 11.568 revistas, mientras que la unión entre las tres bases de datos, es decir el número total de revistas diferentes que se han recuperado entre Ulrich's, WoS, y Scopus ($Ulrich's \cup WoS \cup Scopus$) es de 62.633 revistas.

4.4. Análisis de la cobertura de las bases de datos: geográfica, lingüística, editorial y temática

Cobertura geográfica

Respecto a la cobertura geográfica, en la tabla 7 se observa que en las tres bases de datos los países con más revistas son Estados Unidos y el Reino Unido. En los siguientes puestos ya empieza a haber discrepancias entre las bases de datos. El tercer puesto en Ulrich's lo tiene China, mientras que en WoS y Scopus pertenece a los Países Bajos.

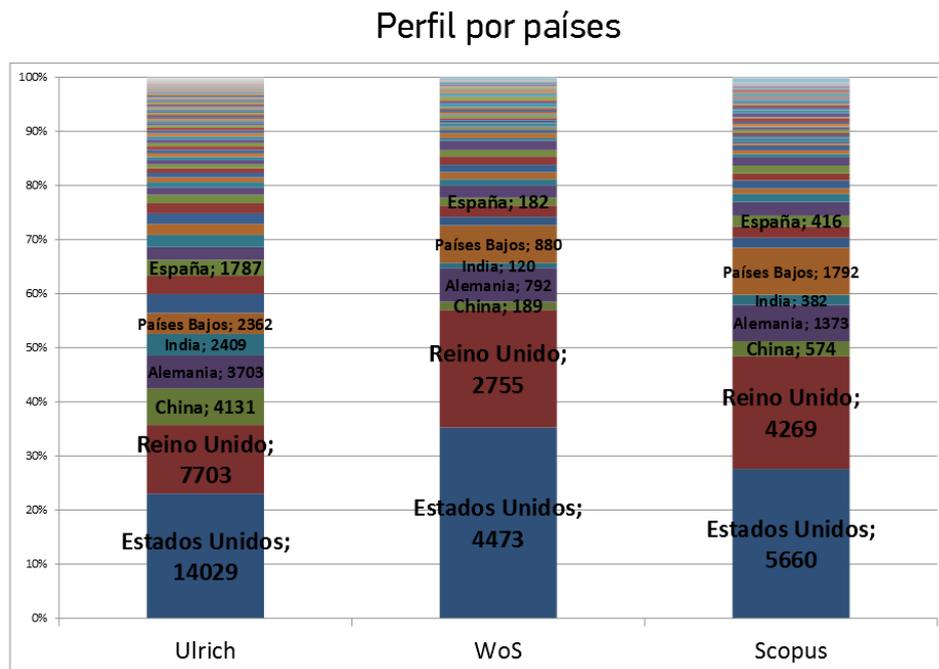
Tabla 7. Los 20 países en los que se editan más revistas en Ulrich's, WoS, y Scopus

| Países | Ulrich | WoS | Scopus |
|----------------|--------|------|--------|
| Estados Unidos | 14029 | 4473 | 5660 |
| Reino Unido | 7703 | 2755 | 4269 |
| China | 4131 | 189 | 574 |
| Alemania | 3703 | 792 | 1373 |
| India | 2409 | 120 | 382 |
| Países Bajos | 2362 | 880 | 1792 |
| Italia | 2129 | 200 | 371 |
| Japón | 2048 | 262 | 429 |
| España | 1787 | 182 | 416 |
| Francia | 1456 | 292 | 513 |
| Brasil | 1360 | 146 | 308 |
| Rusia | 1228 | 163 | 207 |
| Canadá | 1187 | 176 | 303 |
| Australia | 1136 | 191 | 263 |
| Polonia | 974 | 152 | 297 |
| Suiza | 818 | 222 | 313 |
| Rumanía | 601 | 62 | 129 |
| Corea del Sur | 529 | 113 | 136 |
| Turquía | 528 | 74 | 162 |
| Argentina | 484 | 20 | 52 |

En la gráfica 3 se puede observar que Ulrich's contiene una mucha mayor variedad de países (un total de 165) que WoS y Scopus (con 84 y 101 respectivamente). En las tres bases de datos, la mayor proporción de revistas se encuentra en los países anglosajones (Estados Unidos y Reino Unido). Sin embargo, los porcentajes varían mucho entre Ulrich's (donde estos países suman aproximadamente el 35% del total de revistas) y las otras dos bases de datos, pues en WoS el porcentaje asciende a aproximadamente un 58%, y en Scopus casi un 50%. En

el caso de España, Ulrich's es la base de datos donde está mejor representado (3%), frente a WoS (1,5%), y Scopus (2%).

Gráfica 3. Perfil de las bases de datos por países



Cobertura lingüística

Respecto a la cobertura por lenguas, la tabla 8 muestra que el idioma más usado en las revistas es el inglés, con 40609 revistas en Ulrich's, 10607 en WoS, y 16935 en Scopus. En el listado de fuentes de Scopus no aparece la información sobre la lengua en la que están escritos los artículos de las revistas, así que se cruzaron los datos con Ulrich para conseguir esta información.

Tabla 8. Los 15 idiomas más utilizados en las revistas de Ulrich's, WoS, y Scopus

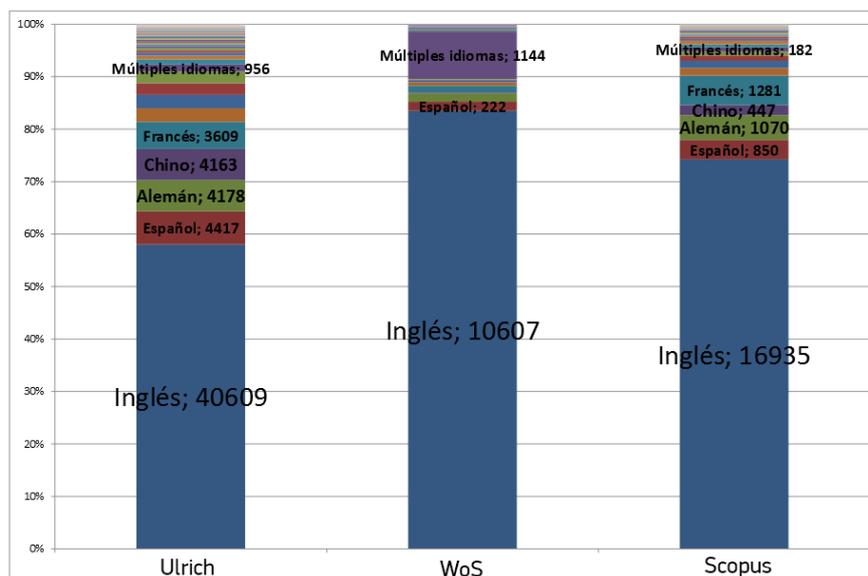
| Lenguas | Ulrich | WoS | Scopus |
|-------------------|--------|-------|--------|
| Inglés | 40609 | 10607 | 16935 |
| Español | 4417 | 222 | 850 |
| Alemán | 4178 | 204 | 1070 |
| Chino | 4163 | 22 | 447 |
| Francés | 3609 | 160 | 1281 |
| Portugués | 1803 | 86 | 320 |
| Italiano | 1796 | 50 | 317 |
| Japonés | 1566 | 12 | 232 |
| Ruso | 1543 | 20 | 172 |
| Múltiples idiomas | 956 | 1144 | 182 |
| Polaco | 683 | 30 | 135 |
| Holandés | 494 | 23 | 152 |
| Turco | 367 | 37 | 86 |
| Checo | 309 | 13 | 85 |
| Ucraniano | 299 | 0 | 16 |

El número total de lenguas que aparecen en las revistas de cada base de datos es también mucho mayor en Ulrich's, que tiene revistas escritas en 118 idiomas diferentes, mientras que en WoS solo se encuentran 30 idiomas, y en Scopus 60.

Es en la gráfica 4 donde se ve más claramente el dominio del inglés sobre todos los demás idiomas. Supone desde un 60% del total de las revistas (en Ulrich), hasta más de un 80% en las revistas de WoS, pasando por el 75% que tiene en Scopus. También es de destacar que el español, aun ocupando aproximadamente solo la décima parte que el inglés, es el segundo idioma más utilizado en las revistas que se indizan en Ulrich's. En WoS, sin embargo, ocupa el tercer puesto (tras “Múltiples idiomas”), y en Scopus, el cuarto (tras el francés y el alemán).

Gráfica 4. Cobertura lingüística de las tres bases de datos

Perfil por lenguas



Cobertura editorial

Respecto a la cobertura editorial de las tres bases de datos, como se observa en la tabla 9, las editoriales con un mayor porcentaje de revistas son las mismas en los primeros puestos, aunque el porcentaje que representan respecto del total son muy diferentes. Así, en Ulrich, las editoriales con más revistas son Elsevier y Springer, con un 3,5% y un 3% del total respectivamente. En WoS también son estas las editoriales con más revistas, pero los porcentajes ascienden a un 12,3% para Elsevier y un 10% para Ulrich. En Scopus la primera editorial también es Elsevier, con un 8,75% del total, pero la segunda es Wiley, con 5,3%, quedando Springer en el tercer puesto con un 4%. De las 25 editoriales que aparecen para cada base de datos en la tabla, hay 15 que aparecen en las 3 bases de datos, y 20 editoriales que están al menos en dos de los Top 25 de cada producto. Además de las editoriales presentadas en esta tabla, en Ulrich hay aproximadamente 28.500 editoriales más, en WoS 3200 y en Scopus, 7200, aunque estos datos no han sido normalizados así que el número puede ser menor.

Tabla 9. Top 25 de editoriales en Ulrich's, WoS, y Scopus

| # | Ulrich's | | | WoS | | | Scopus | | |
|----|----------------------------------|--------------|---------------|---------------------------------|-------------|---------------|---------------------------------|--------------|---------------|
| | Editorial | Fr. Abs. | Fr. Rel. | Editorial | Fr. Abs. | Fr. Rel. | Editorial | Fr. Abs. | Fr. Rel. |
| 1 | Elsevier | 2161 | 3,53% | Elsevier | 1574 | 12,37% | Elsevier | 1792 | 8,75% |
| 2 | Springer | 1897 | 3,10% | Springer | 1289 | 10,13% | Wiley | 1093 | 5,34% |
| 3 | Wiley | 1486 | 2,43% | Wiley | 1241 | 9,75% | Springer | 817 | 3,99% |
| 4 | Routledge | 1102 | 1,80% | Routledge | 432 | 3,39% | Sage Publications | 496 | 2,42% |
| 5 | Sage Publications | 715 | 1,17% | Sage Publications | 427 | 3,35% | Taylor & Francis | 479 | 2,34% |
| 6 | Taylor & Francis | 505 | 0,83% | Taylor & Francis | 365 | 2,87% | Kluwer Academic Publishers | 471 | 2,30% |
| 7 | Inderscience Publishers | 365 | 0,60% | Oxford University Press | 228 | 1,79% | Routledge | 274 | 1,34% |
| 8 | Hindawi Publishing Corporation | 361 | 0,59% | Cambridge University Press | 223 | 1,75% | Lippincott Williams & Wilkins | 236 | 1,15% |
| 9 | Bentham Science Publishers | 358 | 0,59% | Lippincott Williams & Wilkins | 199 | 1,56% | Oxford University Press | 214 | 1,04% |
| 10 | Emerald | 345 | 0,56% | BioMed Central | 160 | 1,26% | BioMed Central | 207 | 1,01% |
| 11 | Oxford University Press | 300 | 0,49% | Informa Healthcare | 148 | 1,16% | Inderscience Publishers | 186 | 0,91% |
| 12 | Cambridge University Press | 294 | 0,48% | IEEE Institute of Electric... | 141 | 1,11% | Cambridge University Press | 168 | 0,82% |
| 13 | Lippincott Williams & Wilkins | 271 | 0,44% | Nature Publishing Group | 96 | 0,75% | Emerald | 163 | 0,80% |
| 14 | BioMed Central | 254 | 0,42% | Karger | 86 | 0,68% | Bentham Science Publishers | 153 | 0,75% |
| 15 | Walter de Gruyter | 229 | 0,37% | Maney Publishing | 70 | 0,55% | IEEE Institute of Electric... | 135 | 0,66% |
| 16 | Informa Healthcare | 206 | 0,34% | Walter de Gruyter | 69 | 0,54% | Hindawi Publishing Corporation | 132 | 0,64% |
| 17 | IEEE Institute of Electric... | 197 | 0,32% | World Scientific Publishing Co | 61 | 0,48% | Haworth Press | 123 | 0,60% |
| 18 | Scientific Research Publishing | 193 | 0,32% | Emerald | 58 | 0,46% | Carfax Publishing | 120 | 0,59% |
| 19 | Research India Publications | 180 | 0,29% | IOP Publishing | 52 | 0,41% | Brill Academic Publishers | 92 | 0,45% |
| 20 | Medknow Publications | 169 | 0,28% | John Hopkins University Press | 50 | 0,39% | Karger | 91 | 0,44% |
| 21 | Omics Publishing Group | 146 | 0,24% | Mary Ann Liebert | 49 | 0,38% | Nature Publishing Group | 80 | 0,39% |
| 22 | I G I Global | 145 | 0,24% | American Chemical Society | 47 | 0,37% | Medknow Publications | 70 | 0,34% |
| 23 | Brill Academic Publishers | 143 | 0,23% | Brill Academic Publishers | 43 | 0,34% | IOS Press | 69 | 0,34% |
| 24 | Karger | 140 | 0,23% | Hindawi Publishing Corporation | 42 | 0,33% | Marcel Dekker | 69 | 0,34% |
| 25 | Izdatel'stvo Nauka | 135 | 0,22% | Georg Thieme Verlag | 41 | 0,32% | Mary Ann Liebert | 62 | 0,30% |
| 26 | Otras editoriales (28560) | 48246 | 78,91% | Otras editoriales (3200) | 5346 | 42,00% | Otras editoriales (7228) | 12413 | 60,60% |

Cobertura temática

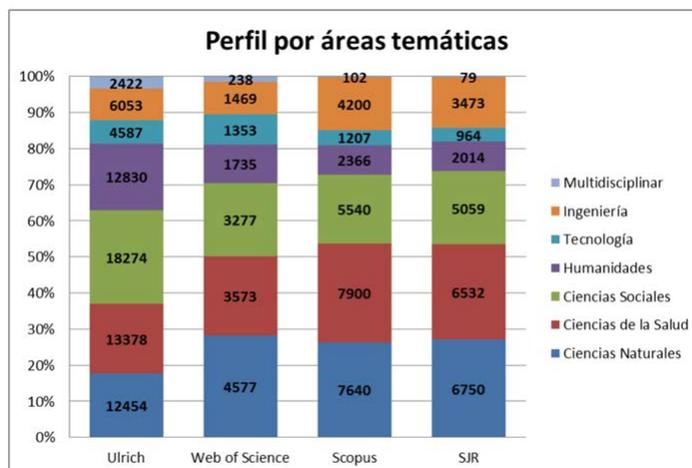
Para agrupar las revistas en una misma lista de áreas temáticas, se han utilizado los sistemas de clasificación de cada producto, y se han asignado sus categorías a seis grandes áreas (Ciencias Naturales, Ciencias de la Salud, Ciencias Sociales, Humanidades, Tecnología, Ingeniería), más otra área donde se han englobado las multidisciplinares. Las asignaciones de cada revista se pueden ver en el Anexo 2 (hoja de cálculo). Tras realizar estas asignaciones, se calculó el número de revistas que caían en cada una de estas grandes áreas. Esto se puede ver en la tabla 10.

Tabla 10. Revistas agrupadas por grandes áreas temáticas

| Áreas | Número de revistas | | | |
|-----------------------------|--------------------|----------------|--------|------|
| | Ulrich | Web of Science | Scopus | SJR |
| Ciencias Naturales | 12454 | 4577 | 7640 | 6750 |
| Ciencias de la Salud | 13378 | 3573 | 7900 | 6532 |
| Ciencias Sociales | 18274 | 3277 | 5540 | 5059 |
| Humanidades | 12830 | 1735 | 2366 | 2014 |
| Tecnología | 4587 | 1353 | 1207 | 964 |
| Ingeniería | 6053 | 1469 | 4200 | 3473 |
| Multidisciplinar | 2422 | 238 | 102 | 79 |

En la gráfica 5 se ve el tamaño proporcional de cada área respecto de la base de datos. En Ulrich's, las revistas de Ciencias Naturales y de la Salud ocupan alrededor de un 35% de la base de datos, mientras que Ciencias Sociales y Humanas ocupa alrededor de un 45% del total. El restante 20%, está ocupado por las revistas de ingeniería, tecnología, y las multidisciplinares. Estas últimas ocupan una proporción mayor en Ulrich's, que en los demás de los productos analizados.

Gráfica 5. Perfil de las bases de datos, por grandes áreas temáticas



En las bases de datos de citas (WoS, Scopus), los porcentajes difieren en gran medida a aquellos encontrados en Ulrich's, pues alrededor del 50% de las revistas pertenecen a las áreas de Ciencias Naturales y Ciencias de la Salud, mientras que alrededor del 30% están incluidas en las categorías de Ciencias Sociales y Humanidades. El 20% restante está constituido por las revistas de Ingeniería, Tecnología, y las multidisciplinarias. En Scopus la cobertura de Ingeniería es superior a la que esta área tiene en WoS (alrededor de un 15% en Scopus, frente a menos del 10% en WoS).

La cobertura de Scopus y SJR es prácticamente la misma, ya que estamos hablando de prácticamente las mismas revistas, y además SJR utiliza el mismo sistema de clasificación que Scopus. Sin embargo, aunque a nivel de categorías se ha detectado que las asignaciones son diferentes a las que hace Scopus, ya que estas asignaciones se han realizado mediante análisis de referencias, como se explica en Gómez-Núñez *et al.* (2011), a nivel de grandes áreas temáticas, estas diferencias no se aprecian demasiado y por eso los resultados que aparecen en la gráfica 5 son prácticamente iguales. En la hoja "Categorías Scopus – SJR" del Anexo 2 se pueden ver las diferencias en las asignaciones de las categorías entre Scopus y SJR a través del número de revistas que tiene asignada cada categoría en cada listado.

5. Conclusiones

En este trabajo se ha diseñado e implementado una *Master List* de revistas científicas a partir del directorio Ulrich's, y de las bases de datos de citas WoS y Scopus (complementadas con los índices bibliométricos JCR y SJR, derivados de las mismas).

Una vez creada la base de datos se ha hecho un estudio sobre control de los ISSN y los títulos, el solapamiento que existe entre las mismas, y los perfiles temáticos, y geográficos (por países y lenguas) que existen en estas bases de datos. Las conclusiones más importantes que se han extraído son:

- El ISSN-L no es un medio eficaz por sí solo para realizar el *matching* entre diferentes versiones del ISSN de una misma revista, ya que por sí solo consigue peores resultados que la unión entre el ISSN y el título proporcionados por la base de datos ($ISSN \cup Tit$), y solo proporciona unas decenas más de emparejamientos que no se podrían haber conseguido de otra manera. Además, hay casos en los que un ISSN no se encuentra en la tabla de correspondencias ISSN/ISSN-L: 8% del total de revistas activas en Ulrich's, 4% en Scopus, y 1% en WoS.
- Thomson Reuters no es transparente en lo que se refiere a identificar qué revistas siguen activas en sus bases de datos, y cuáles no. Scopus, sin

embargo, ofrece esta información de forma gratuita y muy fácilmente accesible.

- El control del ISSN en las revistas es generalmente bueno, siendo este campo uno por los que se consiguen más emparejamientos en el momento de la integración de las bases de datos. Sin embargo, todavía existen revistas que no tienen ningún ISSN asignado en sus respectivos registros: 5,5% en Ulrich's, 2,3% en Scopus, y en el caso de WoS, solo un 0,2%.
- En el estudio del control de los títulos entre las bases de datos se ha encontrado que alrededor de un 30% de las revistas que se habían unificado tras los procesos de *matching* tenían títulos distintos. Se ha calculado la distancia de Levenshtein y un índice de similaridad normalizado para saber en qué proporción una cadena de texto era parecida a la otra, y se ha obtenido que en todos los casos, más de la mitad de las revistas con título diferente tenían una similaridad de al menos el 75%, lo que quiere decir que en la mayor parte de estas revistas con títulos distintos, las diferencias entre los mismos, aunque las hay, son pocas.
- En cuanto al estudio de solapamiento, se han calculado los indicadores TO (traditional overlap) y RO (relative overlap) para cada par de bases de datos (WoS y Scopus, WoS y Ulrich's, Scopus y Ulrich's). Se ha visto que la mayor parte de las revistas de WoS están incluidas en Scopus (96%), y a su vez, que la mayor parte de Scopus está incluida en Ulrich's (94%).
- En la cobertura por países se pone de manifiesto que Estados Unidos y Reino Unido son los países donde más revistas científicas se publican. No obstante, el predominio anglosajón se ve sobrerrepresentado en las bases de datos WoS y Scopus, que detentan porcentajes bastante parecidos.
- En el estudio de cobertura por lenguas, es inapelable el dominio que ejerce el inglés sobre el mundo de las revistas científicas. En Ulrich's, que es donde el porcentaje de revistas en inglés es menor, éste asciende a un 60%, mientras que en WoS, supera el 80%, siguiéndole de cerca Scopus con un 75%.
- En la cobertura de editoriales, hemos visto que en las tres bases de datos, las mismas editoriales ocupan las primeras posiciones, aunque en términos proporcionales, en Ulrich's estas editoriales ocupan un porcentaje mucho menor que WoS y Scopus. También hay mucha más dispersión en Ulrich's que en las dos restantes.
- En los estudios de perfiles temáticos se ha comprobado que en las bases de datos de citas WoS y Scopus tienen cobertura mayoritaria las revistas de Ciencias Naturales y Ciencias de la Salud (~45% mientras que las Ciencias Sociales y las Humanidades ocupan un espacio menor (30%). En el directorio Ulrich's la distribución es diferente (Ciencias Naturales y de la Salud: 35%; Ciencias Sociales y Humanas: 40%).

Algunos desarrollos que se podrían hacer en el futuro tomando como base este trabajo son:

- Aumentar el número de bases de datos integradas en la *Master List*. Se podría ampliar la cobertura, añadiendo más productos bibliométricos como los descritos en la introducción, o también bases de datos especializadas que recogieran alguna información que no estuviera contemplada en las que hay hasta el momento.
- Profundizar en algunos aspectos del control bibliográfico, sobre todo en el control de títulos.

6. Referencias bibliográficas

- Archambault, E., Campbell, D., Gingras, Y., & Lariviere, V. (2009). Comparing of science bibliometric statistics obtained from the web and scopus. *Journal of the American Society for Information Science and Technology*, 60(7), 1320-1326. doi:10.1002/asi.21062
- Archambault, É., Vignola-Gagne, É., Côté, G., Larivière, V., & Gingrasb, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, 68(3), 329-342. doi:10.1007/s11192-006-0115-z
- Braun, T., Glänzel, W., & Schubert, A. (2000). How balanced is the Science Citation Index's journal coverage? A preliminary overview of macrolevel statistical data. En Cronin, B., Atkins, H.B. (Eds), *The Web of Knowledge – A Festschrift in Honor of Eugene Garfield*. Canada: American Society of Information Science, pp. 251-277.
- Gómez-Núñez, A., Vargas-Quesada, B., Moya-Anegón, F., Glänzel, W. (2011). Improving SCImago Journal & Country Rank (SJR) subject classification through reference analysis. *Scientometrics*, 89, pp. 741-758 doi: 10.1007/s11192-011-0485-8
- Goodman, D., & Deis, L.F. (2005). Web of Science (2004 version) and Scopus. *The Charleston Advisor*, 6(3), 5-21. Recuperado el 29 de enero de 2014 desde <http://www.ingentaconnect.com/content/charleston/chadv/2005/00000006/00000003/art00005>
- Goodman, D. y Deis, L. (2007). Update on Scopus and Web of Science. *The Charleston Advisor*, 7(3), p. 15-18. Recuperado el 30 de enero de 2014 desde <http://www.ingentaconnect.com/content/charleston/chadv/2007/00000008/00000003/art00007>
- Delgado López-Cózar, E. (1996). *Las revistas científicas españolas de ciencias de la salud en 1993*. Recuperado el 31 de enero de 2014 desde <http://eprints.rclis.org/13918/>
- Delgado-Lopez-Cozar, E., & Cabezas-Clavijo, A. (2013). Ranking journals: could Google Scholar Metrics be an alternative to Journal Citation Reports and Scimago Journal Rank?. *Learned Publishing*, 26(2), 101-113. doi:10.1087/20130206

- Delgado López-Cózar, E., & Cabezas-Clavijo, A. (2012). Google Scholar Metrics: an unreliable tool for assessing scientific journals. *El Profesional de la Información*, 21(4), 419-427.
doi: [10.3145/epi.2012.jul.15](https://doi.org/10.3145/epi.2012.jul.15)
- Delgado López-Cózar, E., Robinson-García, N., & Torres-Salinas, D. (2013). The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators. *Journal of the American Society for Information Science and Technology*.
doi: [10.1002/asi.23056](https://doi.org/10.1002/asi.23056)
- Delgado López-Cózar, E., & Ruiz-Pérez, R. (2009). La comunicación y edición científica: fundamentos conceptuales. En C. García-Caro & J. Vílchez-Pardo (coord.), *Homenaje a Isabel de Torres Ramírez: Estudios de documentación dedicados a su memoria* (pp. 131-150). Granada: Universidad de Granada.
- Delgado López-Cózar, E., Ruiz-Pérez, R., & Jiménez-Contreras, E. (2006). *La edición de revistas científicas. Directrices, criterios y modelos de evaluación*. Granada: Fundación Española para la Ciencia y la Tecnología.
- Delgado-López-Cózar, E., Jiménez-Contreras, E., Ruiz-Pérez, R., López-Herrera, A.G., Gacto-Colorado, M.J., Torres-Salinas, D., Moneda-Corrochano, M., Ruiz-Baños, R., Pérez-Ortega, J.M., Bailón-Moreno, R., Poyatos-Huertas, E., & Rodríguez-Gálvez, M.J., (2005). In-Recs: Índice de impacto de las revistas españolas de ciencias sociales. Una nueva herramienta para medir el impacto de la investigación española. *Biblio 3W, Revista bibliográfica de geografía y ciencias sociales*, 10(574). Recuperado el 29 de enero de 2014 desde <http://www.ub.es/geocrit/b3w-574.htm>
- Elsevier. (2013). *Scopus content coverage guide*. Recuperado el 27 de enero de 2014, desde: http://cdn.elsevier.com/assets/pdf_file/0019/148402/contentcoverageguide-jan-2013.pdf
- Escalona Fernández, M^a.I., Lagar Barbosa, P., & Pulgarían Guerrero, A. (2010). Web of Science VS. Scopus: un estudio cuantitativo en ingeniería química. *Anales de Documentación*, 13, pp. 159-175. Accedido el 31 de enero de 2014 desde <http://revistas.um.es/analesdoc/article/view/107121/101801>
- Falagas, M. E., Kouranos, V. D., Arencibia-Jorge, R., & Karageorgopoulos, D. E. (2008). Comparison of SClmago journal rank indicator with journal impact factor. *Faseb Journal*, 22(8), 2623-2628.
doi: [10.1096/fj.08-107938](https://doi.org/10.1096/fj.08-107938)
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *Faseb Journal*, 22(2), 338-342.
doi: [10.1096/fj.07-9492LSF](https://doi.org/10.1096/fj.07-9492LSF)
- Fingerman, S. (2006). Web of Science and Scopus: current features and capabilities. *Issues in Science and Technology Librarianship*, (48), 4.
doi: [10.5062/F4G44N7B](https://doi.org/10.5062/F4G44N7B)

- Franceschet, M. (2010). A comparison of bibliometric indicators for computer science scholars and journals on web of science and google scholar. *Scientometrics*, 83(1), 243-258.
doi:[10.1007/s11192-009-0021-2](https://doi.org/10.1007/s11192-009-0021-2)
- Garfield, E. (2007). The evolution of the Science Citation Index. *International Microbiology*, 10, 65-69.
doi: [10.2436/20.1501.01.10](https://doi.org/10.2436/20.1501.01.10)
- Gavel, Y., & Iselid, L. (2008). Web of Science and Scopus: a journal title overlap study. *Online information review*, 32(1), 8-21.
doi:[10.1108/14684520810865958](https://doi.org/10.1108/14684520810865958)
- Gluck, M (1990). A review of journal coverage overlap with an extensión to the definition of overlap. *Journal of the American Society for Information Science*, 41(1), pp. 43-60.
doi: [10.1002/\(SICI\)1097-4571\(199001\)41:1<43::AID-AS14>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-4571(199001)41:1<43::AID-AS14>3.0.CO;2-P)
- Guerrero-Bote, V.P., & Moya-Anegón, F. (2012). A further step forward in measuring journals' prestige: The SJR2 indicator. *Journal of Informetrics*, 6, 674-688.
doi: [10.1016/j.joi.2012.07.001](https://doi.org/10.1016/j.joi.2012.07.001)
- Hood, W.W. & Wilson, C.S. (2003). Overlap in bibliographic databases. *Journal of the American Society for Information Science and Technology*, 54(12) , pp.1091-1103.
doi: [10.1002/asi.10301](https://doi.org/10.1002/asi.10301)
- Jacso, P. (2005). As we may search-Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537.
- LaGuardia, C. (2005), E-views and reviews: Scopus vs. Web of Science. *Library Journal*, 130(1), 40-40. Recuperado el 29 de enero de 2014 desde <http://lj.libraryjournal.com/2005/01/ljarchives/e-views-and-reviews-scopus-vs-web-of-science/>
- Leydesdorff, L., de Moya-Anegón, F., & Guerrero-Bote, V. P. (2010). Journal Maps on the Basis of Scopus Data: A comparison with the Journal Citation Reports of the ISI. *Journal of the American Society for Information Science and Technology*, 61(2), 352-369.
doi:[10.1002/asi.21250](https://doi.org/10.1002/asi.21250)
- Lopez-Illescas, C., de Moya-Anegon, F., & Moed, H. F. (2008). Coverage and citation impact of oncological journals in the web of science and scopus. *Journal of Informetrics*, 2(4), 304-316.
doi:[10.1016/j.joi.2008.08.001](https://doi.org/10.1016/j.joi.2008.08.001)
- Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F. J., González-Molina, A., & Herrero-Solana, V. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, 73(1), 53-78.
doi:[10.1007/s11192-007-1681-4](https://doi.org/10.1007/s11192-007-1681-4)
- Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the web of science for coverage of the social sciences' literature. *Journal of Informetrics*, 1(2), 161-169.

doi:10.1016/j.joi.2006.12.001

ProQuest. *Ulrich's Periodicals Directory*. Recuperado el 27 de enero de 2014, desde <http://ulrichsweb.serialssolutions.com/>

SCImago. (2007). *SJR – SCImago Journal & Country Rank*. Recuperado el 27 de enero de 2014, desde: <http://www.scimagojr.com>

Thomson Reuters. *Journal Citation Reports®*. Recuperado el 27 de enero de 2014, desde <http://webofknowledge.com/>

Thomson Reuters. *Master Journal List*. Recuperado el 27 de enero de 2014, desde <http://ip-science.thomsonreuters.com/mjl/>

Torres-Salinas, D., Bordons, M., Giménez-Toledo, E., Delgado-López-Cózar, E., Jiménez-Contreras, E., & Sanz-Casado, E. (2010). Clasificación integrada de revistas científicas (CIRC): propuesta de categorización de las revistas en ciencias sociales y humanas. *El Profesional de la Información*. 19(6), 675-684. doi:10.3145/epi.2010.nov.15

Anexo 1

Tabla 11. Algunas casuísticas que se presentan cuando los títulos de una misma revista no coinciden entre varias bases de datos

| Casuística | Ejemplo | |
|--|---|---|
| | Variante A | Variante B |
| Utilización de ampersand "&" frente a "and" | Journal of Foot & Ankle Surgery | Journal of Foot and Ankle Surgery |
| Utilización u omisión de "The" al comienzo del título | The Korean Journal of Defense Analysis | Korean Journal of Defense Analysis |
| Inversión del orden de elementos del título | Revista del Hospital Psiquiátrico de la Habana | Hospital Psiquiátrico de la Habana. Revista |
| Adición de elementos cualificadores extraños al título | Spine | Spine (Philadelphia, 1976) |
| Uso de siglas | B i D. Textos Universitaris de Biblioteconomia i Documentacio | BiD |
| Uso del subtítulo de la revista | Journal of Physical Chemistry A | The Journal of Physical Chemistry Part A: Molecules, Spectroscopy, Kinetics, Environment and General Theory |
| Títulos en idiomas diferentes | Metrology and Measuring Systems | Metrologia i Systemy Pomiarowe |
| Uso de títulos paralelos | Aviakosmicheskaya i Ekologicheskaya Meditsina | Aviakosmicheskaya i ekologicheskaya meditsina = Aerospace and environmental medicine |

Anexo 2

Asignación manual de áreas temáticas a las revistas de Ulrich's, WoS, Scopus y SJR.

Comparativas entre las asignaciones de revistas en las categorías de Scopus y SJR. (Hoja de cálculo)