Actas de las 2ª Jornadas de Intercambios y Reflexiones acerca de la Investigación en Bibliotecología, La Plata, 27-28 de octubre de 2011. La Plata: Facultad de Humanidades y Ciencias de la Educación de la Universidad Nacional de La Plata, 2011. ISSN: 1853-5631

Las fuentes de datos en los estudios bibliométricos

Claudia E. Boeris[1]

[1]Instituto Argentino de Radioastronomía, CONICET. biblio@iar.unlp.edu.ar

Resumen. Los datos necesarios para llevar a cabo estudios bibliométricos se obtienen de fuentes de distinta naturaleza que, por lo general, no han sido diseñadas con ese propósito. Los aspectos cuantificables de la literatura científica están ligados a elementos de dato presentes en los documentos científicos tales como título, autores, afiliación institucional de los autores, resumen, palabras clave y referencias bibliográficas. Todos los estudios bibliométricos operan sobre alguno de estos elementos en función de sus objetivos y de la disponibilidad de la información necesaria para realizarlos.

Se evalúan tres fuentes de datos con el objetivo de determinar la presencia o ausencia de los elementos mencionados, su normalización y su adecuación para realizar este tipo de estudios.

Introducción y marco teórico

En la actualidad los estudios bibliométricos son una herramienta valiosa para describir las estructuras de investigación a nivel internacional, nacional, regional o institucional y analizar el impacto que estas estructuras tienen en la generación de conocimiento.

Para realizar este tipo de estudios es necesario contar con los datos que se incluyen en las publicaciones científicas y que a su vez se encuentran representados en bases de datos referenciales o en catálogos de bibliotecas. Los datos a seleccionar dependerán del tipo de análisis que se desee realizar. Así por ejemplo, los nombres de los autores, su afiliación, las palabras clave o las palabras presentes en los títulos, las fechas de publicación o las fechas incluidas en las referencias son algunos de los elementos ampliamente utilizados.

Otro de los aspectos que deben considerarse es el grado de cobertura que poseen las fuentes de datos en relación con la temática y con el periodo temporal a estudiar, como así también la normalización de los términos. (Gavel y Lars, 2008) propone analizar las posibilidades que brinda la interfaz de búsqueda, la cobertura temática, la consistencia

en la indexación de los términos, el uso de un tesauro y las funcionalidades para la recuperación de citas bibliográficas.

Según el tipo de estudio bibliométrico que se desee realizar se deberán tener en cuenta diferentes elementos o aspectos a analizar. Es importante saber cuál es tratamiento que la fuente de datos hace de ellos como así también si están o no presentes en dicha fuente. En la tabla 1 se detallan los tipos de estudio y los aspectos a considerar.

Tabla1. Tipo de estudio y aspectos a analizar

Aspecto a analizar	**Tipo de estudio**
Afiliación	Estudios por institución, por país, estudios de colaboración científica.
Fechas de publicación	Obsolescencia, rangos a estudiar
Autores	Frentes de investigación, colaboración científica
Revistas o libros fuente	Núcleo de publicaciones, frentes de investigación
Referencias y citas	Estudios de impacto, visibilidad, obsolescencia
Palabras clave o descriptores	Frentes de investigación, temáticas
Palabras del título o del resumen	Frentes de investigación, áreas temáticas

Las fuentes de datos disponibles presentan características y funcionalidades de distinta naturaleza y muchas veces sus elementos de dato no están uniformados o normalizados lo cual puede ocasionar distorsiones en los resultados. La evaluación de estas fuentes puede contribuir a la mejora de los servicios que ofrecen y adecuarlas a los requerimientos de los estudios bibliométricos.

Metodología

Se compararon la base de datos Scopus, el NASA Astrophysics Data System (ADS) y la base de trabajos publicados por los investigadores del Instituto Argentino de Radioastronomía (IAR). Scopus y ADS son bases de datos internacionales, Scopus es multidisciplinaria y ADS especializada en astronomía, física, matemáticas e instrumentación. La base del IAR es una base bibliográfica local especializada en radioastronomía.

Para comparar los tres servicios se tuvieron en cuenta la normalización de los términos,

las posibilidades de recuperar cada elemento de dato, la exportación de registros en distintos formatos y las funcionalidades para la recuperación de citas y referencias. Se analizaron también los servicios adicionales que presta cada una de las bases consultadas. Se evaluaron las funcionalidades de cada sistema determinando si las herramientas de recuperación estaban disponibles en la interfaz de usuario o si había que disponer de herramientas auxiliares para la extracción de los datos.

Resultados y discusión

Cada una de las fuentes analizadas presentan características particulares. Las bases de datos referenciales ofrecen herramientas de recuperación que se adecuan más a las necesidades del análisis bibliométrico. Por su parte la base de datos del IAR ofrece instrumentos de recuperación más limitados, y en este caso particular es necesario optimizar la recuperación mediante el uso herramientas que operan a nivel de la propia base, como por ejemplo la aplicación de los utilitarios CISIS o la flexibilidad que ofrece el formato MARC 21 en los campos de longitud fija y de códigos.

Los resultados obtenidos se agruparon en cinco categorías de análisis:

1) Normalización de datos
2) Posibilidades de recuperación
3) Posibilidades de exportación
4) Funcionalidades para la recuperación de citas y referencias
5) Otros servicios

En cada uno de los puntos se evaluó la disponibilidad de herramientas de recuperación en la interfaz de usuario o bien la necesidad de contar con instrumentos adicionales de procesamiento.

1) Normalización de datos: en este contexto se entiende por normalización la existencia de una forma unívoca para las entradas en la base de datos. Los resultados se muestran en la Tabla 2.

Tabla 2. Normalización de datos

Autores	
Scopus	Las entradas no están normalizadas, sin embargo se establece la relación entre las diferentes formas del nombre de una persona

	utilizando los datos de afiliación
ADS	Las entradas no están normalizadas. Se pueden recuperar las diferentes formas del nombre a partir de un índice de autores
IAR	Las entradas están normalizadas. En el caso de haber diferencias en las formas del nombre no se establecen relaciones.
Afiliación	
Scopus	Las entradas están normalizadas y se presentan las variantes en el registro que representa la institución en la base de datos. (funcionaría como un registro de autoridad)
ADS	Las entradas no están normalizadas. Solo se puede recuperar la afiliación a partir de las palabras presentes en títulos y resúmenes.
IAR	No incluye datos de afiliación
Descriptores y palabras clave	
Scopus	Recuperable desde la interfaz. Búsqueda por palabra clave. No posee tesauro. Posibilidad de recuperar por áreas del conocimiento. Incluye palabras clave dentro de los registros exportados.
ADS	Recuperable desde la interfaz. Búsqueda por palabra clave. No posee tesauro. Incluye palabras clave dentro de los registros exportados.
IAR	Recuperable desde la interfaz. Búsqueda por palabra clave o descriptor. Posee índice temático. Posibilidad de recuperar palabras clave o descriptores en los registros.
Títulos de las revistas fuente	
Scopus	Los títulos están normalizados. Presenta lista de fuentes cubiertas. No posee códigos de título.
ADS	Los títulos están normalizados. Posee códigos que identifican los títulos. Se necesitan herramientas de procesamiento para extraer los datos. Ejemplos:

	JO - Monthly Notices of the Royal Astronomical Society Bibliographic Code: 2007MNRAS.378..947S %G MNRAS journal = {\mnras}, (formato de exportación BibTeX)
Base del IAR	Los títulos de las revistas fuente para los artículos de la base están descriptos en el subcampo 773^t de MARC. No están totalmente normalizados. La interfaz incluye acceso al índice de títulos. Se necesitan herramientas de procesamiento para extraer los datos. El catálogo incluye un campo local para identificar el título mediante un código generado a partir del tipo de fuente.
Fechas	
Scopus	Posee formato de fecha normalizado mediante la exportación de un archivo delimitado por separadores (.csv). Ejemplos: con separadores: 2011 texto: (2011)
ADS	No hay una forma normalizada general para recuperar el año de publicación. La normalización opera a nivel del formato de exportación. Una vez obtenidos los registros se necesitan herramientas de procesamiento. Ejemplos: Y1 - 2010/12/1 Publication Date: 12/2010 %D 12/2010 year = 2010, (formato de exportación BibTeX)
Base del IAR	Las fechas están normalizadas en función de AACR2 y es posible recuperarlas desde la interfaz de usuario y desde el campo 008 de MARC21. Se necesitan herramientas de procesamiento para extraer los datos.

En el caso de la base del IAR los registros creados con AACR2 presentan la dificultad de incorporar elementos de descripción que no permiten recuperar las fechas exactas de

publicación, i. e. [197-?]. El formato MARC a partir del campo 008 soluciona en parte este problema pues las fechas se incluyen en forma numérica eliminándose los corchetes y otros signos, no obstante mantiene algunos caracteres para fechas indefinidas, i.e. 196u.

2) Posibilidades de recuperación. Tabla 3

Recuperar por afiliación	
Scopus	Es posible recuperar por nombre de la institución, por un código único de institución, como así también afiliación por ciudad, por país y por organización.
ADS	Se incluyen los datos de afiliación pero no posee una herramienta de recuperación directa a través de la interfaz. Es posible recuperar las afiliaciones a partir de uno o más registros y éstas se muestran en una tabla ordenada por apellido del autor. Puede exportar la lista a un archivo con delimitadores, a formato Excel y texto.
Base del IAR	No incluye datos.
Recuperar por rangos de fechas	
Scopus	Por rango de uno o más años (publicación o de incorporación en la base de datos).
ADS	Por rango de años, meses o días (publicación o de incorporación en la base de datos).
Catáogo IAR	Rango de fechas por año en la interfaz avanzada.
Identificacion y recuperación de fuentes (revistas)	
Scopus	Por tipo de documento, por título de revista incluyendo palabras del título o título exacto, por por ISSN, por DOI, por volumen.
ADS	Por código de revista, por año/vol/nro o buscar por referencia y por código bibliográfico del ADS que individualiza cada abstract en forma unívoca, y desde el formato etiquetado en la exportación de registros.
Base del IAR	Recuperable por interfaz (índice), y mediante herramientas adicionales de procesamiento

3) Funcionalidades de exportación. Tabla 4

Exportación de registros	
Scopus	Formatos: texto (ASCII), Refworks, BibTeX, RIS y archivo con separadores (procesable mediante una planilla de cálculo)
ADS	Se puede exportar en diferentes formatos. HTML abstracts, texto ASCII, BibTeX, listados, generic tagged abstracts, EndNote, ProCite, Refman, RefWorks, MEDLARS, Dublin Core XML, XML abstracts, XML references, RSS, AASTeX, Icarus, MNRAS, Customizable. Uno de ellos etiquetado en formato de texto (Procite) que permite crear una base de datos de trabajo.
Base del IAR	No posee desde la interfaz, se obtienen mediante herramientas adicionales de procesamiento.
Selección de datos a exportar	
Scopus	Es posible seleccionar el tipo de dato que se quiere exportar: solo citas, solo abstracts, abstracts con referencias, formato completo o bien seleccionar los datos que se quieren exportar.
ADS	El tipo de dato a exportar depende del formato de exportación elegido.
Base del IAR	No posee desde la interfaz, se obtienen mediante herramientas adicionales de procesamiento.

4) Funcionalidades para la recuperación de citas y referencias. Tabla 5

Recuperación de citas	
Scopus	Tiene la posibilidad de realizar análisis de citas: a partir de un conjunto de documentos la muestra ordenándolas por fecha. Muestra el total de citas de los últimos tres años. Es posible eliminar las autocitas, ordenar los documentos y seleccionar por rango de fecha. Es posible recuperar las citas como registros de la propia base Scopus.
ADS	Es posible recuperar las citas como registros de la propia base

	ADS, es posible eliminar las autocitas, el sistema les asigan un peso y ofrece varias opciones de ordenamiento.
Base del IAR	No incluye datos.
Recuperación de referencias	
Scopus	Es posible recuperar las referencias a los documentos como un conjunto de registros de la base Scopus.
ADS	Es posible recuperar las referencias a los documentos como un conjunto de registros de la base ADS.
Base del IAR	No incluye datos.
Aplicación de filtros: Los filtros permiten refinar los resultados de las búsquedas para lograr precisión.	
Scopus	Posee filtros desde la interfaz de búsqueda por año de publicación, autor, área temática, tipo de documento (artículo, review, conferencia, etc.), palabra clave, afiliación, título de la fuente, tipo de fuente (monografía, tesis, reporte, etc.), lengua, ciudad y país.
ADS	Posee filtros por base de datos (astrofísica, física, archivo de preprints), por objeto astronómico, por trabajos con/sin referato, por tipo de instrumento astronómico (este filtro puede permitir generar estudios de producción para una determinada institución), por fecha (día, mes, año), por título de fuente, por elemento dentro del registro (citas, referencias, artículos a texto completo, tabla de contenidos, contenido multimedia, etc.)
Base del IAR	No posee filtros.
Ordenamiento de resultados	
Scopus	Resultados ordenados por fecha, nro. de citas, relevancia, autor, título de la fuente.
ADS	Ofrece diferentes forma de ordenamiento: por fecha, por *score (*cada registro es calificado en función de la relevancia para la búsqueda realizada), por citas y citas normalizadas, por primer, por cantidad de autores, por ingreso en la base de datos. El sistema se asigna un peso a cada documento recuperado de

	acuerdo al ordenamiento elegido.
Base del IAR	Ordenamiento por autor/título, título, fecha, ubicación (signatura topográfica).

5) Otros servicios. Tabla 6

Servicios al usuario	
Scopus	Posee alertas, permite salvar registros, búsquedas. Posee ayuda en línea.
ADS	Es posible crear colecciones de registros. FAQ y ayuda en línea.
Base del IAR	No permite crear perfil de usuario. Posee ayuda en línea
Cálculo del Índice H	
Scopus	Scopus lo calcula y muestra gráfico.
ADS	Ordena los resultados y asigna un peso a cada trabajo incluyendo también la cantidad de citas. No lo calcula. Pero es posible extraerlo en forma manual.
Base del IAR	No incluye datos.

Conclusiones

La normalización de los términos facilita la tarea de selección de datos y asegura la fiabilidad de los resultados. Cada una de las bases analizadas presentan algún tipo de normalización, no obstante no puede decirse que son ciento por ciento satisfactorias.

Con respecto a las funcionalidades de recuperación y exportación Scopus y el ADS ofrecen herramientas de recuperación más accesibles para un usuario final. La base de datos local necesita del conocimiento de los utilitarios CISIS para la extracción de datos, lo cual representa una barrera para usuarios no expertos, sin embargo esta característica permite realizar un diseño en las estrategias de búsqueda que no se limita a las funcionalidades que ofrece una interfaz y por lo tanto es más adaptable a necesidades particulares.

Un elemento relevante para los estudios bibliométricos es la posibilidad de recuperar y analizar citas y referencias, en este sentido se puede decir que Scopus sería la herramienta más satisfactoria siguiéndole el ADS. El catálogo local no sería una buena fuente en este caso.

La elección de la fuente de datos a utilizar en un estudio bibliométrico se basa fundamentalmente en la capacidad de aquélla para satisfacer y cubrir todos los aspectos que se deseen estudiar. Puede decirse que en la actualidad no todas las fuentes poseen todas las funcionalidades deseables para lograr ese objetivo, por tal motivo es recomendable analizar las características de cada una y evaluar el uso conjunto de las herramientas disponibles. Sería deseable que los proveedores de bases de datos se esforzaran por ofrecer mejores herramientas de extracción y procesamiento como así también que se normalicen los datos.

Bibliografía

- Gavel, Ylva y Lars, Iselid. 2008. Web of Science and Scopus: a journal title overlap study. Online Information Review, Vol. 32, No. 1, p. 8-21
- Base de datos de trabajos publicados del IAR [en línea]. [Citado 06 Sep 2011]. Disponible en World Wide Web: http://www.iar.unlp.edu.ar/biblio/cgi-bin/opacmarc/wxis?IsisScript=opac/xis/opac.xis&db=contri
- NASA Astrophysics Data System [en línea]. [Citado 06 Sep 2011]. Disponible en World Wide Web: http://www.adsabs.harvard.edu/
- Scopus [en línea]. [Citado 06 Sep 2011]. Disponible en World Wide Web: http://www.scopus.com/home.url

Bibliografía Consultada

- Astro2010: The Astronomy and Astrophysics Decadal Survey, Position Papers, nª 28 [en línea]. [Citado 06 Sep 2011]. Disponible en World Wide Web: http://adsabs.harvard.edu/abs/2009astro2010P.28K
- Carlstein, Stefan. Bibliometrics. University library, Jönköping University [en línea]. [Citado 06 Sep 2011]. Disponible en World Wide Web: http://hj.se/bibl/en/publishing/bibliometrics.html
- Kurtz, Michael J., Eichhorn, Guenther, Accomazzi, Alberto. 1999. The NASA ADS Abstract Service and the distributed Astronomy Digital Library. D-Lib Magazine, vol. 5, nª 11 [en línea]. [Citado 06 Sep 2011]. Disponible en World Wide Web: http://www.dlib.org/dlib/november99/11kurtz.html

- Kurtz, Michael, Accomazzi, Alberto, Murray, Stephen. 2009. The Smithsonian/NASA Astrophysics Data System (ADS) Decennial Report [en línea]. [Citado 06 Sep 2011]. Disponible en World Wide Web: http://adsabs.harvard.edu/abs/2009astro2010P..28K
- Lee, Jonghoon y Dubin, David. 2003. Vocabulary Mapping in the NASA ADS: prospects for practical subject access. En: Brenda Corbin et al., editores. Library and Information Services in Astronomy [en línea], vol. 4, p. 249. [Citado 06 Sep 2011]. Disponible en World Wide Web: http://adsabs.harvard.edu/abs/2003lisa.conf.249L
- Peter's Digital Reference Shelf. Scopus revisited [en línea]. [Citado 05 Sep 2011]. Disponible en World Wide Web: http://www.jacso.info/gale/Scopus-revisited/bbb/scopus-revisited.htm
- Vieira, Elizabeth S. y Gomes, José A. N. F. 2009. A comparison of Scopus and Web of Science for a typical university. Scientometrics. vol. 81, nª 2, p. 587-600.

Software utilizado:

- awk, grep, sort: comandos del sistema operativo Linux que posibilitan extraer, buscar y ordenar datos presentes en archivos de computadora.
- Bibexcel: conversión y tratamiento de datos bibliográficos [en línea]. [Citado 02 Sep 2011]. Disponible en World Wide Web: http://www8.umu.se/inforsk/Bibexcel/
- CiteSpace: análisis, visualización y agrupación de datos bibliográficos obtenidos de Web of Science [en línea]. [Citado 02 Sep 2011]. Disponible en World Wide Web: http://cluster.cis.drexel.edu/~cchen/citespace/
- FUSE: software para el procesamiento de registros del ADS [en línea]. [Citado 02 Sep 2011]. Disponible en World Wide Web: http://www.eso.org/sci/libraries/telbib+FUSE.html
- GNUplot: utilitario para la representación gráfica de los datos [en línea]. [Citado 02 Sep 2011]. Disponible en World Wide Web: http://www.gnuplot.info/
- Publish or Perish: Se basa en los datos de Google Scholar. Calcula el valor de diferentes indicadores como el Índice H, y realiza análisis bibliométrico [en línea]. [Citado 02 Sep 2011]. Disponible en World Wide Web:

http://www.harzing.com/pop.htm

- Pajek: software para visualizar y analizar redes. Bibexcel puede ser usado para generar las matrices que necesita Pajek [en línea]. [Citado 02 Sep 2011]. Disponible en World Wide Web: http://pajek.imfm.si/doku.php
- R: lenguaje de programación usado para análisis estadísticos y construcción de gráficos [en línea]. [Citado 01 Sep 2011]. Disponible en World Wide Web: http://www.r-project.org/
- SITKIS: software multipropósito. Es compatible con la exportación de redes [en línea]. [Citado 01 Sep 2011]. Disponible en World Wide Web: http://users.tkk.fi/~hschildt/sitkis/index.html
- Utilitarios CISIS: permiten manipular las bases de datos CDS/ISIS a nivel de la línea de comandos y aplicando el lenguaje de formateo de CDS/ISIS [en línea]. [Citado 01 Sep 2011]. Disponible en World Wide Web: http://bvsmodelo.bvsalud.org/php/level.php?lang=es&component=28&item=1