

BIBLIOMÉTRIA: HERRAMIENTA PARA LA IDENTIFICACIÓN, DISTRIBUCIÓN Y EVOLUCIÓN DE LA LITERATURA CIENTÍFICA

Héctor PÉREZ MORAGO
Dr. en Ingeniería Informática
Profesor-tutor UNED Calatayud

Gema MERINO ANGULO
Lda. Historia del Arte
Graduada en Magisterio

Resumen: Una de las consecuencias de la expansión de internet es la posibilidad de acceder a una mayor cantidad de información. Este hecho no es ajeno a la comunidad científica que ha pasado a disponer de una gran cantidad de información de forma casi inmediata. Sin embargo, la capacidad humana para poder interpretarla, analizarla y generar nuevo conocimiento a partir de ella, no ha aumentado al mismo ritmo. De aquí surge la Bibliometría con objeto de, mediante la aplicación práctica de modelos matemáticos, identificar el tamaño, distribución, estructura social y evolución de la literatura científica. Su aplicación práctica es un proceso complejo compuesto de varios pasos que requiere del desarrollo de herramientas software que asistan a los usuarios. El objetivo de este trabajo consiste en describir que es la Bibliometría, mostrar el flujo desarrollo de un análisis bibliométrico y revisar una serie de herramientas desarrolladas específicamente para realizar este tipo de análisis.

Palabras clave: Bibliometría; mapa científico; análisis de rendimiento herramientas bibliométricas.

Abstract: One of the consequences of the expansion of the Internet is the possibility of accessing a greater amount of data. This fact is not alien to the scientific community, which has become available to a large amount of data almost immediately. However, the human capacity to be able to interpret, analyze and generate new knowledge from it, has not been increased at the same pace. This is where Bibliometric arises in order, through the practical application of mathematical models, to identify the size, distribution, the social structure and evolution of the scientific literature. The practical application of Bibliometric is a complex process composed of several steps that requires the development of software tools that assist users. The objective of this work is to describe what Bibliometric is, to provide a methodology to carry out a bibliometric analysis, and to review a series of tools developed specifically to carry out this kind of analysis.

Keywords: Bibliometrics; Science Mapping; Performance Analysis, Bibliometrics tools.

1. INTRODUCCIÓN

Una de las consecuencias de la extraordinaria expansión de internet es la posibilidad de acceder a una cada vez mayor cantidad información. Este hecho también se aplica en el caso de la literatura científica, que aparte de ser cada vez más accesible (de acuerdo con OASPA el número de artículos publicados de acceso abierto para el periodo 2010-2018 fue de 1,392,051 de los cuales 261,621, es decir más del 23%, fueron publicados en 2018¹) también ha aumentado en su volumen. De acuerdo con (Edge, 1979) y (Solla Price, Solla Price, Solla Price, & Solla Price, 1986) el número de publicaciones científicas ha experimentado una tasa crecimiento exponencial en los últimos cincuenta años.

En este contexto, los recopiladores de datos de publicación científica como Scopus², Web of Science³ (WoS), google scholar⁴ (GS) o Dimensions ⁵han pasado a ser herramientas imprescindibles para cualquier persona interesada en el mundo de la investigación. Sin embargo, y a pesar de que dichas bases ayudan al acceso e identificación de todo tipo de publicaciones científicas, se precisan de nuevas técnicas que permitan recuperar y extraer la información relevante de forma automática con el fin de descubrir nuevas líneas de investigación, patrones, tendencias, relaciones entre grupos de investigación, etc.

En los últimos años se han desarrollado, en diversas áreas de investigación como la inteligencia artificial o la minería de textos, nuevas técnicas para el tratamiento de grandes cantidades de información. Entre estas técnicas se encuentran los sistemas de recuperación de información, los sistemas de recomendaciones o los sistemas de clasificación de documentos. En el caso de la literatura científica, el área de investigación que se ocupa de tal fin es la *bibliometría*. La bibliometría aplica métodos matemáticos y estadísticos a la literatura científica con el objetivo de estudiar y analizar la actividad científica (Pritchard & others, 1969). A este proceso de análisis y estudio se le conoce como análisis bibliométrico y generalmente, se lleva a cabo mediante técnicas como: *mapa científico* (science mapping) y *análisis de rendimiento* (performance analysis) (Cobo M. J., 2012).

Los mapas científicos, mapas bibliométricos o cienciogramas son un tipo de mapa conceptual que trata de representar de forma espacial los campos, disciplinas, especialidades o publicaciones, las relaciones que existen entre estos elementos, así como la evolución de los mismos (Small, 1973).

Por otro lado, la técnica conocida como análisis de rendimiento, trata de evaluar de forma cuantitativa el rendimiento de un área utilizando para ello indicadores bibliométricos y unidades de análisis como pueden ser las instituciones, los autores o las fuentes de publicación, entre otras (Garfield, 1977).

1. Datos extraídos de <https://oaspa.org/growth-continues-for-oaspa-member-oa-content/>

2. <https://www.scopus.com/home.uri>

3. <https://www.webofscience.com/wos/woscc/basic-search>

4. <https://scholar.google.es/>

5. <https://www.dimensions.ai/>

Llevar a la práctica un análisis bibliométrico requiere de herramientas software que faciliten el trabajo con grandes cantidades de información. Algunas de estas herramientas son proporcionadas desde los propios recopiladores de publicaciones científicas como es el caso de WoS (InCites⁶) o Scopus (SciVal⁷). Dichos recopiladores de publicaciones incorporan herramientas para el análisis de citas o para la visualización de la distribución de documentos en el tiempo. También existen herramientas software de propósito general como Microsoft Excel, librerías para lenguajes de programación como bibliometrix o BiblioTools (para los lenguajes de programación R y Python, respectivamente), paquetes de software estadístico como SPSS o herramientas para el análisis de redes sociales como Pajek (Batagelj & Mrvar, 1998). Sin embargo, debido a que el análisis bibliométrico es un proceso complejo compuesto de varias etapas, en los últimos años, se han desarrollado herramientas software específicas que asisten de diversa forma al proceso de análisis bibliométrico. Algunas de las más utilizadas son: Bibexcel (Persson, Danell, & Schneider, 2009), Bibliometrix (Aria & Cuccurullo, 2017), CitNetExplorer (Van Eck & Waltman, CitNetExplorer: A new software tool for analyzing and visualizing citation networks, 2014), SciMAT (Cobo M. J., 2012) o VOSViewer (Van Eck & Waltman, Software survey: VOSviewer, a computer program for bibliometric mapping, 2010).

El objetivo de este trabajo es mostrar que es la bibliometría, mostrar una metodología para realizar análisis bibliométricos, así como dar una serie de herramientas diseñadas para asistir al usuario en dicho proceso. En este trabajo nos vamos a centrar en aquellas herramientas software que han sido diseñadas para asistir en el proceso del análisis bibliométrico. En concreto, vamos a analizar las siguientes herramientas bibliométricas, por ser las más utilizadas: Bibexcel, Bibliometrix, CitNetExplorer, SciMAT y VOSViewer. Para cada una de ellas, mostramos cuáles son sus principales características, así como sus principales ventajas e inconvenientes.

Como veremos más adelante, cada una de estas herramientas poseen características particulares que las hacen más adecuadas en función del tipo de análisis, destacando que la mayoría de ellas son complementarias, y que, por tanto, pueden utilizarse conjuntamente para facilitar el proceso de análisis.

Lo resta de este trabajo se encuentra organizado del siguiente modo: en la Sección 2 metodología revisamos cual es proceso habitual para realizar un análisis bibliométrico. La Sección 3 describe en detalle dos de las técnicas más utilizadas para llevar a cabo un análisis bibliométrico: el análisis de rendimiento y los mapas científicos. A continuación, en la Sección 4 presentamos cada de las herramientas bajo estudio, proporcionando para cada una de ellas sus principales características, así como sus principales ventajas e inconvenientes. Finalmente, este trabajo termina en la Sección 5 mostrando las principales conclusiones extraídas de este trabajo.

6. <https://incites.clarivate.com/>

7. <https://www.scival.com/>

2. METODOLOGÍA

En esta sección mostramos de forma genérica cuales son los principales pasos o etapas de las que se compone el flujo de trabajo de un análisis bibliométrico. En concreto, vamos a seguir la metodología propuesta en (Börner, Chen, & Boyack, 2003) y (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011) que se compone de los siguientes pasos: selección de las fuentes de datos, construcción de la consulta a las fuentes de datos, preprocesamiento de los datos extraídos, construcción de las redes de análisis, análisis y por último interpretación del resultado del análisis. La Figura 1 muestra de forma visual los diferentes pasos de los que consta el análisis bibliométrico.

En lo que resta de esta sección revisamos cada uno de estos pasos con mayor detalle.



Figura 1. Flujo de trabajo

1. **Selección** de la o las **fuentes de datos** que utilizaremos en el análisis. En la actualidad existen diversas fuentes de datos bibliográficos que nos permiten tanto la búsqueda como la recuperación de información de las principales áreas de investigación. La Tabla 1 muestra para cada de las fuentes de datos más utilizadas, su nombre, el enlace de acceso al recurso, y las áreas de publicación que recogen.

Nombre	url	Áreas
WoS	https://www.webofknowledge.com	Multidisciplinar
Scopus	https://www.scopus.com/	Multidisciplinar
GS	https://scholar.google.com/	Multidisciplinar
Dimensions	https://www.dimensions.ai/	Multidisciplinar
NLM's MEDLINE	https://www.ncbi.nlm.nih.gov/pubmed	Biomédica
arXiv	https://arxiv.org/	Algunas áreas
CiteSeerX	http://citeseerx.ist.psu.edu/	Informática
Digital Bibliography & Library Project	https://dblp.uni-trier.de/	Informática
SAO/NASA Astrophysics Data System	http://adswww.harvard.edu/	Astronomía Física
Science Direct	https://www.sciencedirect.com/	Algunas áreas

Tabla 1. Fuente de datos

Conviene aclarar que existen tanto recopiladores que solo se ocupan de una materia o disciplina concreta como recopiladores multidisciplinares. Sin embargo, un aspecto para tener en cuenta, en relación a dichos recopiladores multidisciplinares, es que no cubren del mismo modo todas las áreas de investigación (Yang & Meho, 2006). En este punto conviene subrayar que la calidad del análisis realizado depende fuertemente de la calidad del análisis de los datos trabajados. Por tanto, es de crucial importancia seleccionar como origen de datos una fuente que sea al mismo tiempo, fiable y que cubra el mayor número de registros posible. En la actualidad, los siguientes recopiladores multidisciplinares de datos de literatura científica son las más utilizadas:

- Scopus de la editorial Elsevier (Netherlands). Es una de las mayores bases de datos de literatura y citas bibliográficas bajo licencia. Contiene más de 21,900 títulos de más de 5,000 editoriales de todo el mundo, incluyendo revistas científicas, libros y actas de congresos. Lo que permite tener una visión global de la producción de investigación en el mundo en los campos de la ciencia, la tecnología, la medicina, las ciencias sociales, así como de las artes y las humanidades. Además, Scopus integra una serie de herramientas inteligentes para rastrear, analizar y visualizar la investigación como es el caso de SciVal.

- Web of Science (WoS) recientemente adquirido por Clarative analytics. Al igual que Scopus es una de las principales bases de datos de literatura científica bajo licencia. Entre sus principales características se encuentran: más de un billón citas, captura e indexa citas desde 1900, indexa revistas, libros, actas de congresos, etc. Fue la primera en incluir herramientas de análisis bibliométrico (Journal Citation Reports (JCR) primera herramienta para el análisis y clasificación de las revistas científicas y profesionales, e InCites primera herramienta de evaluación que permite el análisis basado en citas de las instituciones de investigación).

- Google scholar herramienta de la compañía Google de acceso libre que permite buscar e identificar material bibliográfico publicado de carácter científico-académico. Cubre todo tipo de documentos como libros, artículos de revistas científicas, comunicaciones y ponencias a congresos, informes científico-técnicos, tesis, tesinas o memorias de grado, trabajos científicos depositados en repositorios, archivos de preprints, páginas web personales o institucionales, etc. En general, podemos afirmar que cubre cualquier publicación con resumen. Su mayor crítica es que no existe ningún tipo de control en los procesos de selección de los documentos que indexa (Halevi, Moed, & Bar-Ilan, 2017).

- Dimensions es producto fruto de la colaboración entre seis empresas del portafolio de Digital Science (Altmetric, Digital Science Consultancy, Figshare, Readcube, Symplectic y UberResearch), más de 100 investigadores y universidades. su objetivo es conseguir que la investigación sea más accesible. Incorpora varios servicios, siendo algunos de ellos de pago. Entre sus principales características podemos mencionar: cuenta con más de 117 millones de documentos, y más de 1,4 billones de citas, proporciona una API para realizar consultas utilizando un DSL (Domain Specific Language) similar a SQL (Structured Query Language) que pueden ser lanzadas desde casi cualquier lenguaje de programación. El resultado de dichas consultas son conjuntos de registros en formato JSON (Java Simple Object Notation) que facilitan su procesamiento y análisis.

2. Una vez seleccionado el origen de los datos a obtener, es necesario definir de forma precisa la **consulta** a lanzar sobre dichas fuentes. En general, todas las fuentes de datos bibliográficas soportan diferentes operadores como pueden ser operadores lógicos (and, or, not) para realizar consultas, elementos auxiliares (paréntesis, comillas), así como diferentes parámetros para refinar las mismas (selección de un periodo temporal, tipo de documento, fuentes de publicación, etc.).

Como ejemplo, la Figura 2 muestra la consulta "(binary decision diagrams or ordered reduced binary decision diagrams or ORBDD or ROBDD)" en WoS. El resultado de la consulta arroja un total de 2,931 documentos.

Estos documentos suponen el punto de partida del análisis, pero antes de poder llevarlo a cabo se requiere de una etapa de preprocesamiento, la cual pasamos a comentar en el siguiente punto.



Figura 2. Consulta WoS

3. Una vez obtenidos los datos objeto de estudio es necesario realizar una fase de **preprocesamiento**. Esta es una de las etapas más importantes del análisis ya que los datos recuperados podrían contener errores, imprecisiones, omisiones, etc. Por

ejemplo, es habitual encontrar errores lexicográficos, términos escritos en singular y plural, acrónimos, palabras clave (keywords) diferentes que representan un mismo concepto, documentos a los que le falta la institución, etc. En este punto es precioso recalcar nuevamente que la calidad de los resultados del análisis vendrá determinada por la calidad de los datos objeto de estudio.

4. **Construcción** de redes de análisis. Este paso a su vez puede descomponerse en tres pasos: extracción de los elementos de análisis, normalización y construcción de la red. Para la extracción de los elementos de análisis debemos seleccionar en primer lugar cual es el tipo de elementos que queremos analizar. Existen en la literatura relacionada diferentes propuestas atendiendo a las diferentes unidades que podemos utilizar en el análisis (autores, documentos, palabras clave, etc.). Entre las más utilizadas podemos destacar: *análisis de términos comunes* (co-words analysis) (Callon, Courtial, & Laville, 1991), que tiene por objetivo descubrir la estructura conceptual de un campo de investigación utilizando para ello determinadas palabras representativas de los documentos (palabras clave, palabras extraídas de los abstract y títulos, etc.); *análisis de autores comunes* (co-author analysis) (Peters & Van Raan, 1991) y (Glänzel, 2001), que permite identificar colaboraciones entre instituciones o investigadores; *análisis de referencias comunes* (bibliographic coupling analysis) (Kessler, 1963), que tiene por objeto identificar documentos que tratan sobre materias relacionadas analizando para ello las referencias que tienen en común (como podemos ver en la Fig. 3.b los documentos a y b citan al mismo tiempo a los documentos c y d); *análisis de citas comunes* (co-citation analysis) (Small, 1973), que actúa de forma similar al análisis de referencias comunes. En este caso, se dice que dos documentos son cocitados si existe al menos un documento que cita a ambos (como podemos ver en la Fig. 3.a los documentos c y d son citados al mismo tiempo por el documento a y también por el b).

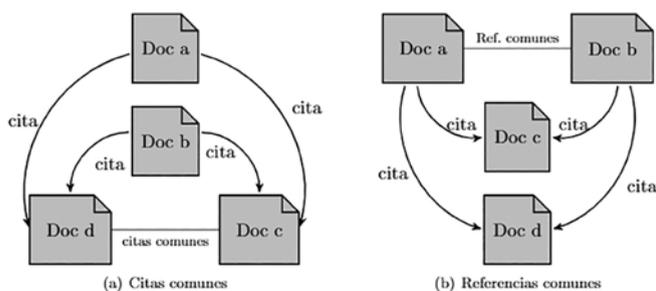


Figura 3. Análisis

Una vez se han extraído los elementos de análisis, hemos seleccionado el tipo de análisis y construido las redes bibliométricas, es necesario realizar un proceso de normalización con el fin de obtener similitudes entre sus nodos. Este proceso de normalización nos permite evitar algunos problemas como por ejemplo la aparición de términos poco frecuentes que aparecen juntos casi siempre. En la práctica llevar a cabo a este proceso requiere de la elección de alguna medida de similitud.

Se han propuesto diferentes medidas de similitud siendo las más utilizadas Coseno de Salton, Índice de Jaccard, Fuerza de asociación (también conocida como Índice de proximidad o Índice de afinidad probabilística). Emplazamos al lector a (Eck & Waltman, 2009) donde se muestra un estudio comparativo de las diferentes medidas de similitud.

5. **Análisis.** En esta etapa llevamos a cabo el análisis bibliométrico propiamente dicho. Hay dos técnicas bibliométricas ampliamente utilizadas y que son complementarias: análisis de rendimiento y mapas científicos. Dichas técnicas son revisadas en detalle en la Secciones 3.1 y 3.2, respectivamente.

6. **Interpretación** de los resultados del análisis. En esta última fase los resultados del análisis de los datos objeto de estudio tienen que ser interpretados para extraer conclusiones que den respuesta a los objetivos marcados al inicio del estudio. En esta última fase se requiere de la participación de un experto en el área para poder contrastar los resultados obtenidos en el análisis.

3. ANÁLISIS

En esta sección revisamos en detalle dos de las principales técnicas bibliométricas: mapas científicos y análisis de rendimiento.

3.1 Análisis de Mapas científicos.

Los mapas científicos o mapas bibliométricos son un tipo de mapa conceptual que representan de forma espacial los campos, disciplinas, especialidades o publicaciones, así como las relaciones que existen entre estos elementos (Small, 1973). De este modo, los mapas se convierten en herramientas potencialmente útiles para los usuarios ya que permiten descubrir información relevante acerca de la estructura y dinamismo de un campo científico.

Existen dos tipos de elementos, que son comunes a la mayoría de los diagramas independientemente de la herramienta utilizada, que son: los *ítems* y los *enlaces*.

Los ítems pueden representar autores, instituciones, documentos, fuentes de publicación, países, palabras clave, etc.

Por otro lado, los enlaces representan relaciones o conexiones entre esos ítems. Como ejemplo, un enlace entre dos autores puede representar una relación de coautoría. Además, en relación a su representación gráfica, cabe destacar que el tamaño de estos elementos es proporcional al peso que tienen en la muestra objeto de estudio para el caso de los ítems, y a la fortaleza de la relación para el caso de los enlaces. Es decir, cuanta mayor es la relación entre dos ítems mayor es el grosor de la línea que los une. Un aspecto para tener en cuenta es que no suele ser habitual mezclar elementos de distinto tipo en un mismo diagrama.

Es posible establecer una clasificación de mapas científicos atendiendo al fin para el que fueron construidos. De este modo podemos clasificar los mapas en:

- *Mapas de colaboración.* Este tipo de mapas permiten visualizar las relaciones existentes entre diferentes actores como por ejemplo autores, grupos de investigación, instituciones, países, etc. De forma genérica podemos decir que estos son construidos

utilizando para ello: las referencias, las citas entre publicaciones, la coautoría de publicaciones, etc. Por ejemplo, utilizando las afiliaciones de los autores podemos extraer cual es el grado de internacionalización de una determinada área de investigación.

• *Mapas de tendencias.* Este tipo de mapas permiten visualizar patrones o tendencias de investigación dentro de un área o ámbito bajo estudio como puede ser el ámbito de publicación de una revista, de un grupo de investigación, de un país o de una institución, entre otros. Como ejemplo podemos citar: los mapas temáticos, los mapas estratégicos, los mapas de evolución conceptual, etc. De forma genérica podemos señalar que estos son construidos utilizando para ello: palabras o palabras clave extraídas del objeto de investigación. Dichas palabras pueden ser proporcionadas por el autor, por la revista, por el editor, o por el creador de la base de datos de donde se ha extraído los registros bajo estudio.

3.2 Análisis de rendimiento.

Existen diferentes propuestas para evaluar el rendimiento de un área de investigación generalmente basadas en indicadores bibliométricos como: el número de publicaciones, porcentajes de trabajos indexados en alguna fuente de publicación o el número de citas recibidas, entre otros.

Una de las técnicas más utilizadas para evaluar el rendimiento de un área de investigación se basa en el análisis de citas, que parte de la siguiente premisa:

“un artículo es más relevante o tiene un mayor impacto cuanto mayor es el número de citas que ha recibido”.

Sin embargo, utilizar las citas recibidas de un documento de forma directa presenta los siguientes inconvenientes: por un lado, el número de citas promedio entre los diferentes campos de investigación es variable. Por otro lado, determinados tipos de documentos científicos, como los de revisión literaria, presentan un mayor número de citas promedio. Para solventar estos inconvenientes, se han propuesto diferentes indicadores como h-index (Hirsch, 2005), g-index (Egghe, 2006) o hg-index (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009). De estos indicadores, uno de los más utilizados es el h-index, que puede ser definido como:

“Un científico tiene un índice h si y solo si tiene n documentos con h o más citas, y el resto de sus documentos tienen h o menos citas”.

Utilizando el h-index junto con el concepto “citation classics”, introducido por (Garfield, 1977), podemos identificar aquellos documentos denominados clásicos dentro de un campo de investigación sujeto a estudio. En concreto, utilizando la adaptación del h-index propuesta por (Martínez, Herrera, López-Gijón, & Herrera-Viedma, 2014) podemos identificar los h documentos de un área científica que tienen h o más citas, teniendo el resto de los documentos h o menos citas. Este conjunto de documentos es conocido como el H-core del área, y entre sus principales características podemos destacar que nos permite identificar los mayores avances en el área, los principales actores del área (autores, instituciones, países, etc.) u observar su evolución desde una perspectiva histórica.

4. HERRAMIENTAS BIBLIOMÉTRICAS

Como se ha comentado en la introducción de este trabajo existen varias herramientas para llevar a cabo un análisis bibliométrico, las cuales pueden ser clasificadas atendiendo al fin para el que han sido desarrolladas. Es decir, herramientas incluidas en las propias bases de datos de literatura científica, herramientas de propósito general y herramientas específicamente creadas para el análisis bibliométrico. En este trabajo nos vamos a centrar en aquellas herramientas específicamente creadas para el análisis bibliométrico. En concreto, las herramientas tratadas en el presente trabajo son las que se listan en la Tabla 2.

Herramienta	Autor	Institución
Bibexcel ¹	Olle Persson	Univ. de Umeå (Suecia)
bibliometrix ²	M. Aria y C. Cuccurullo	Univ. de Napoles Federico II (Italia)
Citespace ³	Chaomei Chen	Univ. de Drexel (EE.UU.)
SciMAT ⁴	M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma y F. Herrera	Univ. de Granada (España)
VOSviewer ⁵	Nees Jan van Eck y Ludo Waltman	Univ. de Leiden (Holanda)

Tabla 2. Herramientas bibliométricas

4.1 Bibexcel.

Bibexcel es una herramienta free-ware para uso académico desarrollada por Olle Persson (Universidad de Umeå, Suecia). Su diseño está adaptado para asistir a los usuarios en el análisis de datos bibliográficos, o cualquier otro tipo de datos textuales formateados de una manera similar. La idea es generar ficheros de datos que puedan ser importados a Microsoft Excel, SPSS IBM o cualquier otro programa que acepte como entrada datos tabulares, para su posterior procesamiento. Entre sus características principales podemos destacar:

- Es un programa versátil que funciona extrayendo campos de las fuentes de datos proporcionadas.
- Soporta registros de WoS y Scopus. Además, permite convertir otros formatos como USPTO, WinSpirs, Dialog, etc.
- Analiza las frecuencias de términos y construye las relaciones entre estos para generar las matrices y vectores que representan ese tipo de análisis.
- Permite realizar análisis de términos comunes, de autores comunes, de referencias comunes y de citas comunes. También permite construir matrices de co-ocurrencias de cualquier campo del documento o combinaciones de estos.
- Incorpora herramientas para la fase de preprocesamiento extraer lexemas, transformación de texto o encontrar documentos duplicados.
- Este software requiere de otro software para proporcionar salidas gráficas. Por ejemplo, podemos utilizar Pajek o SPSS [3] para mostrar de forma gráfica los resultados obtenidos.

4.2 Bibliometrix.

Bibliometrix es un software openSource para el lenguaje de programación R desarrollado por Massimo Aria y Corrado Cuccurullo de la universidad de Nápoles. Puede ser utilizado directamente en R importando la librería o mediante una interfaz web denominada Biblioshiny que proporciona una forma de trabajo para perfiles no desarrolladores. Entre sus principales características podemos destacar:

- Integración con otros paquetes estadísticos y gráficos del lenguaje R.
- Gran comunidad de usuarios lo que permite rápidas actualizaciones.
- Soporte para importar datos de Socopus, WoS, Cochrane Library, Dimensions y Pub Med.
- Filtrado de datos.
- Soporte para el Análisis y gráficos para tres niveles diferentes de métricas: fuente, autores y documentos.
- Análisis de tres estructuras de conocimiento: estructura conceptual, estructura intelectual y estructura social.
- Contiene varias herramientas que permiten realizar un análisis bibliométrico completo. Una de sus principales características es que permite construir muchos tipos diferentes de gráficos, característica esta que no está presente en todas las herramientas analizadas.

4.3 Citespace.

CiteSpace es una herramienta analítica visual, desarrollada en Java, para encontrar tendencias y patrones en la literatura científica. La herramienta ha sido desarrollada por Chaomei Chen (Universidad de Drexel (EEUU)). Su diseño se basa en la visualización progresiva del dominio del conocimiento, mediante la búsqueda de puntos críticos en el desarrollo del campo o dominio. Sus características principales se pueden resumir en:

- Proporciona diversas funciones para facilitar la comprensión e interpretación de los patrones de red y patrones históricos, incluyendo la identificación de las áreas de rápido crecimiento temático, la búsqueda de “puntos calientes” de citación de las publicaciones, la descomposición de una red en clusters, patrones geospaciales de colaboración y áreas de colaboración internacional.
- Soporta análisis estructurales y temporales de una variedad de redes derivadas de publicaciones científicas. Por ejemplo, redes de colaboración, redes de citas comunes de autores, de documentos, de referencias comunes, de términos comunes, etc. También admite redes de tipos de nodos híbridos tales como términos, instituciones y países, y tipos de enlaces híbridos como citas comunes, términos comunes y enlaces de citas dirigidas.
- Soporta varias fuentes de datos como WoS, PubMed, arXiv, ADS o NSF Award Abstracts, entre otros.
- Se puede utilizar para generar superposiciones de mapas geográficos visibles en Google Earth en función de la ubicación de los autores.

4.4 SciMAT.

SciMAT es una herramienta open source bajo licencia (GPLv3) que asiste al usuario en el análisis de mapas científicos bajo un marco longitudinal. Se compone de varios módulos cada uno diseñado para facilitar el trabajo en cada una de las etapas del análisis bibliométrico: uno dedicado a la gestión de la base de conocimiento, otro dedicado a realizar el análisis de mapa científico, y otro dedicado a visualizar los resultados generados en el análisis. Entre sus principales características se encuentran:

- Soporte para importar datos en los formatos WoS y RIS.
- Permite construir redes bibliométricas de términos comunes, de citas comunes entre autores, de citas comunes entre revistas, de referencias comunes, etc.
- Soporte la etapa de preprocesamiento, con funcionalidades como eliminación de duplicados (de forma automática y manual), agrupación de términos, agrupación por períodos de tiempo, reducción de datos y reducción de redes.
- Soporta las siguientes medidas de normalización: Coseno de Saltón, Índice de Jaccard, Fuerza de asociación, índice de equivalencia.
- Soporta los siguientes tipos de análisis: de redes, de rendimiento y de calidad, y análisis longitudinal.
- Permite la visualización de las distintas redes generadas: diagramas estratégicos, mapas temáticos, mapas de evolución, etc.
- Los informes pueden ser obtenidos en formato HTML y LaTeX.

4.5 VOSviewer.

VOSviewer es una herramienta software para construir y visualizar redes bibliométricas. Ha sido desarrollada por el CWTS (Centre for Science and Technology Studies) de la universidad de Leiden. Sus redes de análisis pueden ser construidas utilizando relaciones de citas, referencias comunes, citas comunes o de coautoría utilizando como unidad de análisis fuentes de publicación (revistas, congresos, etc), investigadores o artículos. Entre sus principales características se encuentran:

- VOSviewer permite también la construcción y visualización de redes de términos comunes extraídos del corpus a analizar, utilizando funcionalidad de minería de texto.
- Como fuente de datos acepta registros de WoS, Scopus, Dimensions, PubMed y en formato RIS. Además, permite, mediante el uso de API Rest incorporar datos de Crossref, Europe PubMed Central, Semantic Scholar, Open citations, Microsoft Academic (MA) y WikiData.
- Incorpora herramientas que evitan el solapamiento de las etiquetas en los gráficos generados.
- Permite exportar sus salidas gráficas en varios formatos lo que dota a esta herramienta de una gran flexibilidad.
- La herramienta permite ser ejecuta en su versión de escritorio o se lanzada desde su página web.

4.6 Comparativa.

La Tabla 3 muestra de forma resumida las principales características de las herramientas analizadas en las subsecciones previas.

Herramienta	Fuente de datos	Preprocesamiento	Gráficas
Bibexcel	WoS, Scopus, ProCite, Bibexcel	Documentos duplicados	Requiere de un software externo
bibliometrix	WoS, Dimensions, Scopus	Filtros y división por periodos de tiempos	Si
Citespace	WoS, Dimensions, Scopus, MA, Crossref, PubMed, ADS, arXiv, CNKI, Cscsi, Derwent Patents, NSF, ProQuest, Fulltext, CSV	Filtros y división por periodos de tiempos	Si
SciMAT	WoS, Scopus, CSV	Filtros, división por periodos de tiempos, distancia entre palabras, detecta singulares/plurales, edición de los datos	Si
VOSviewer	WoS, Dimensions, Scopus, MA, Crossref, RIS, PMC, Semantic Scholar, Open Citation, WikiData	Filtros	Si

5. CONCLUSIONES

En este trabajo hemos presentado que es la bibliometría. Asimismo, se han descrito dos de sus principales técnicas de análisis: mapas científicos y el análisis de rendimiento. Además, hemos mostrado una metodología, identificando sus principales pasos, para llevar a cabo con éxito un análisis bibliométrico.

Por otro lado, en la segunda parte de este trabajo se han analizado 5 herramientas que nos permiten automatizar el proceso de análisis bibliométrico. Como resultado de este análisis hemos comprobado que ninguna de las herramientas analizadas es superior en todas sus características al resto. Por el contrario, cada una de ellas posee una serie características particulares que la hace más adecuada en función del tipo de análisis a realizar. En general, podemos afirmar que las herramientas bibliométricas pueden ser utilizadas de forma complementaria, aprovechando sus particularidades, y, por tanto, mejorando la calidad del análisis realizado.

BIBLIOGRAFÍA

Alonso, S., Cabrerizo, F., Herrera-Viedma, E., & Herrera, F. (2009). hg-index: A new index to characterize the scientific output of researchers based on the h-and g-indices. *Scientometrics*, *82*, 391-400.

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*, 959-975. doi:https://doi.org/10.1016/j.joi.2017.08.007

Batagelj, V., & Mrvar, A. (1998). Pajek-program for large network analysis. *Connections*, *21*, 47-57.

Bonacich, P. (2008). Exploratory social network analysis with Pajek. SAGE PUBLICATIONS INC 2455 TELLER RD, THOUSAND OAKS, CA 91320 USA.

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual review of information science and technology*, *37*, 179-255.

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, *22*, 155-205.

Cobo, M. J. (2012). SciMat: herramienta software para el análisis de la evolución del conocimiento científico. *Propuesta de una metodología de evaluación, Granada: Universidad de Granada*.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, *62*, 1382-1402.

Eck, N. J., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, *60*, 1635-1651.

Edge, D. (1979). Quantitative measures of communication in science: A critical review. *History of science*, *17*, 102-134.

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, *69*, 131-152.

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. *The FASEB journal*, *22*, 338-342.

Garfield, E. (1977). Introducing citation classics-human side of scientific reports. INST SCI INFORM INC 3501 MARKET ST, PHILADELPHIA, PA 19104.

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, *51*, 69-115.

Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature. *Journal of Informetrics*, *11*, 823-834.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, *102*, 16569-16572.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American documentation*, *14*, 10-25.

Martínez, M. A., Herrera, M., López-Gijón, J., & Herrera-Viedma, E. (2014). H-Classics: Characterizing the concept of citation classics through H-index. *Scientometrics*, *98*, 1971-1983.

Persson, O., Danell, R., & Schneider, J. W. (2009). How to use Bibexcel for various types of bibliometric analysis. *Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday*, *5*, 9-24.

Peters, H., & Van Raan, A. (1991). Structuring scientific activities by co-author analysis: An exercise on a university faculty level. *Scientometrics*, 20, 235-255.

Pritchard, A., & others. (1969). Statistical bibliography or bibliometrics. *Journal of documentation*, 25, 348-349.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for information Science*, 24, 265-269.

Solla Price, D. J., Solla Price, D. J., Solla Price, D. J., & Solla Price, D. J. (1986). *Little science, big science... and beyond*. Columbia University Press New York.

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *scientometrics*, 84, 523-538.

Van Eck, N. J., & Waltman, L. (2014). CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of informetrics*, 8, 802-823.

Yang, K., & Meho, L. I. (2006). Citation analysis: a comparison of Google Scholar, Scopus, and Web of Science. *Proceedings of the American Society for information science and technology*, 43, 1-15.

LISTADO DE FIGURAS

Figura 1. Flujo de trabajo, página 4.

Figura 2. Consulta WoS, página 7.

Figura 3. Análisis, página 8.

Las figuras 1 y 3 han sido generadas en overleaf mediante el uso de los siguientes paquetes para el lenguaje de procesamiento LATEX: smartdiagram y TikZ.

La figura 2 es una captura de la página web *Web of Science* realizada el día 20 de marzo de 2021.

(Footnotes)

1 <https://homepage.univie.ac.at/juan.gorraiz/bibexcel/>

2 <http://www.bibliometrix.org/>

3 <http://cluster.cis.drexel.edu/~cchen/citespace/>

4 <https://sci2s.ugr.es/scimat/>

5 <https://www.vosviewer.com/>