

Herramientas de análisis de datos bibliográficos y construcción de mapas de conocimiento: Bibexcel y Pajek

[Versió catalana]

ALICIA GARCÍA-GARCÍA, ALBERTO PARDO-IBÁÑEZ
Universitat Catòlica de València
alicia.garcia@ucv.es, alberto.pardo@ucv.es
ANTONIA FERRER SAPENA, FERNANDA PESET
Universitat Politècnica de València
anfersa@upv.es, mpeasetm@upv.es
LUIS M. GONZÁLEZ-MORENO
Universitat de València
luis.m.gonzalez@uv.es

Resumen

Los mapas de conocimiento son instrumentos que ayudan a visualizar información. Si toman como base del análisis datos de naturaleza bibliográfica se convierten en una forma sencilla de proporcionar el estado de la cuestión de un área o tema muy concreto. Este trabajo se centra en la utilización de información científica bibliográfica para producir un mapa, presentando las herramientas para construirlos. Bibexcel es utilizado para analizar los datos extraídos de la base de datos Web of Science y Pajek para representar de forma visual los resultados. El trabajo describe los pasos necesarios para utilizar estas aplicaciones, especialmente cuando no son adecuadamente dibujados por contar con un número excesivo de datos.

Resum

Els mapes de coneixement són instruments que ajuden a visualitzar informació. Si prenen com a base de l'anàlisi dades de naturalesa bibliogràfica es converteixen en una forma senzilla de proporcionar l'estat de la qüestió d'una àrea o un tema molt concret. Aquest treball se centra en la utilització d'informació científica bibliogràfica per produir un mapa, i presenta les eines per construir-los. Bibexcel s'usa per analitzar les dades extretes de la base de dades *Web of Science* i Pajek, per representar visualment els resultats. El treball descriu els passos necessaris per utilitzar aquestes aplicacions, especialment quan no mostren un dibuix adequat a causa d'un nombre excessiu de dades.

Abstract

Knowledge maps are tools that help visualize information. If they take bibliographical data as the basis for network analysis they become a simple means of getting to the crux of highly specific areas or subjects. This paper focuses on the use of scientific bibliographic information to produce a knowledge map and describes the tools that can do this. Bibexcel is used to analyze the data extracted from Web of Science and Pajek is used to visually represent the results. The paper describes the various steps in the procedure, focusing on what to do when visualization is hampered by too many data.

Rebut: 25/11/2014. Acceptat: 18/02/2015.

1 Concepto

Los mapas de conocimiento son representaciones gráficas que muestran de forma visual el estado de una cuestión determinada. "The visual representations take advantage of the human eye's broad bandwidth to allow users to understand abstract information in an intuitive way. In this sense, among the most common visual representations are maps" (Pino-Díaz et al., 2012). Para construir estos mapas se necesitan datos, que pueden residir en personas o en fuentes externas.

La representación de datos en mapas de conocimiento procede de las disciplinas que estudian la gestión del conocimiento, es decir, las ligadas a la gestión de empresa y a la gestión de la información como la documentación. En el caso de las empresas "One of the most challenging problems in decision analysis is capturing information that resides in decision-makers and experts. [...] The knowledge we have about any uncertain event is composed of many fragmented pieces of information that are relevant to the event in question. The fragments of information may exist in one person or among several people" (Howard, 1989). Los mapas de conocimiento en este ámbito, por tanto, resultan de gran ayuda para identificar y representar los recursos de conocimiento de los que dispone una organización (Alós, 2003). Entendemos estos recursos como un conjunto disperso de capacidades, competencias, documentos, procedimientos, tecnologías, etc. que necesitamos conocer para poder gestionarlos eficazmente de forma que apoyen la idea de negocio de la organización (González, 2013).

En el campo de la documentación, la gestión del conocimiento se vincula al análisis de la información y también a los instrumentos que se emplean para ello, como los tesauros y las ontologías. "Si los tesauros se proponen controlar el vocabulario y limitar el sentido de los términos para facilitar la recuperación [de la información] por la semejanza entre petición y representación o para sugerir términos de búsqueda [,] los mapas conceptuales presentan la estructura conceptual de una disciplina mediante una representación gráfica de la significación de los conceptos." (Hernández Quintana, 2006).

Dado que la gestión del conocimiento es una disciplina de carácter social, el resultado está determinado por la formación y cultura del sujeto que realiza la acción de describirla o recuperar la información. Esta singularidad, ya demostrada en campos como la etnografía, puede ser superada desde posturas filosóficas que incluyan un cierto relativismo cultural (Sánchez Durá, 2013) o con la ayuda de técnicas objetivas de procesamiento de datos, como es el *text mining*. Por todas estas razones, en el entorno de la gestión de la información la explotación de datos se ha convertido en uno de los grandes protagonistas del análisis estratégico, por ejemplo de los sectores científico técnicos. Mediante técnicas de minería de textos sobre datos bibliográficos se muestra desde el estado de un área de conocimiento, un grupo de organismos o lugares (Herrero y Moya, 2009; Peset et al., 2013; Valenciano et al., 2010; Leydesdorff, 2013; Alexandre et al., 2013; Robinson et al., 2013; Ardanuy, 2014), se generan indicadores científico técnicos (Ruas y Pereira, 2014) o se experimentan con técnicas para conocer las necesidades de los usuarios (Torres Salinas et al., 2014).

Nuestro trabajo se inserta en este contexto, la representación del conocimiento de un sector, con la intención de proporcionar una guía clara para analizar datos procedentes de una fuente bibliográfica y representar sus relaciones en una red de términos. El análisis de datos mediante técnicas de *text mining* y su visualización recientemente se ha convertido en una disciplina consolidada (Tsai, 2012, Nualart-Vilaplana et al., 2014). La denominación del tema que presentamos varía desde el nombre escogido por nosotros, mapas de conocimiento, a redes de colaboración, mapas de la ciencia, grafos, etc.

2 Función

Al inicio de todo proceso de investigación —especialmente si se trata de una área ajena a nuestra experiencia— se realiza una labor de documentación centrada en la búsqueda de la bibliografía adecuada. De ella se extrae el estado de una cuestión, lo que tradicionalmente (Tarrats Pons, 2012) se producía mediante "sucesivas inmersiones del investigador en las fuentes documentales [...]. El proceso de búsqueda consumía mucho tiempo, y era difícil de replicar dados los juicios subjetivos implicados" (Börner; Chen; Boyack, 2003).

Las bases de datos sobre artículos de revista donde se busca esa información de calidad, estructurada y fiable se erigen además en fuentes de datos para otro tipo de análisis. Según Pino et al. (2012), "Swanson (1986, 1987) showed the potential of bibliographic databases for knowledge discovery, in particular by using a text-mining tool of concept linkage. [...] Co-word analysis (Callon, Law, & Rip, 1986; Michelet, 1988; Courtial & Michelet, 1990) is a method used in knowledge discovery in bibliographic databases". Dado el montante de información científica que se produce, las técnicas tradicionales de extracción de conocimiento, como la lectura, están dejando paso a otras formas de análisis (Filippov, 2014). La evolución tecnológica ha ayudado con nuevas formas de enfoque y capacidades para la ciencia, como la colaboración entre grupos de investigación u otras técnicas de análisis de información (López-Borrull; Canals, 2013).

Por ejemplo, el *text mining* puede realizar cálculos matemáticos y estadísticos sobre algunos campos de los artículos para revelar sus patrones o estructura interna, así como la evolución de una determinada disciplina. Esta metodología aumenta la objetividad de las conclusiones que se derivan de la interpretación de estos mapas. El recuento de aparición de los términos, ya sean autores o palabras clave, está siendo utilizado masivamente en las nubes de *tags* gracias a sencillas utilidades en la web. Wordle o Tagxedo presentan una figura en la que con distintos tamaños de letra destacan las etiquetas más repetidas en la colección. Aunque para realizar algún recuento más preciso que indique, por ejemplo, la cantidad de repeticiones han de utilizarse funcionalidades más específicas. Algunas se ofrecen desde las propias bases de datos Web of Science-WoS (Web of Science®, Thomson Reuters, New York, USA) y Scopus (Elsevier Properties, Netherlands), aunque también existen herramientas generadas para ellas como HistCite (HistCite Software LLC, New York, USA) o Bibexcel (Olle Persson, Umeå University, Umeå, SWE).

Frente a los recuentos, más o menos sencillos, un análisis complejo de la forma en que se relacionan los términos, de las co-ocurrencias, co-autoría o *co-word*, arroja conclusiones más interesantes. El estudio de redes y su visualización en forma de mapa o red de relaciones sociales se basa en la teoría de grafos, una rama de las matemáticas. De hecho, "desde el momento en el que un texto es fragmentado en palabras y se establece algún tipo de relación entre dichas palabras, disponemos de una representación en forma de grafo" (Cruz Mata et al., 2006). La representación de una red social también mediante un grafo con nodos (individuos) conectados por medio de líneas (relaciones) aborda otros recuentos, que manifiestan conceptos como el de centralidad, determinada por el grado de conexiones de un nodo con otros (Del-Fresno-García, 2014).

3 Herramientas para generar mapas de conocimiento

Existen varias herramientas para construir redes y representar las relaciones entre términos. *Text2mindmap* (<https://www.text2mindmap.com/>) o *Mindmeister* (<http://www.mindmeister.com/es> (<http://www.mindmeister.com/es>)) son de uso libre y permiten crear mapas más o menos sencillos introduciendo manualmente la información.

Pero construir mapas de conocimiento que provengan de fuentes externas y además tengan en cuenta ciertos indicadores —recuento de frecuencia de aparición junto a las relaciones entre los términos— requiere utilizar herramientas específicas. En el campo de la bibliometría se han generado numerosas aplicaciones (Cobo et al., 2011). En este texto vamos a explicar el funcionamiento básico de dos programas gratuitos: Bibexcel en combinación con Pajek (versión 3.14, de 12 de noviembre de 2013 Batagelj and Mrvar, University of Ljubljana, Ljubljana, Slovenia).

Para ilustrar este trabajo realizaremos una consulta en la Colección principal (i.e. *Core Collection*; <http://acesowok.fecyt.es> (<http://acesowok.fecyt.es/>)) de WoS con el término *Knowledge maps* en el campo "Tema", restringida al área de Ciencias de la Información y Ciencias Bibliotecarias.¹ Los registros obtenidos se guardan con el nombre "savedrecs.txt" en un archivo que constituirá la fuente de datos necesaria para construir el mapa de conocimiento. La opción se encuentra en "Guardar en otros formatos de archivo" y debemos hacerlo con las siguientes especificaciones:

- en contenido del registro: registro completo y referencias citadas para posteriores usos
- en formato del archivo: texto sin formato

Si el resultado de la búsqueda fuera mayor de 500 registros, hay que descargarlos de 500 en 500 y después unirlos en un sólo archivo, ya sea con un programa para editar texto plano (WordPad) o desde Bibexcel (menú "File/Append one file to another").

Bibexcel es un programa versátil que extrae campos, analiza las frecuencias de términos y construye las relaciones de los términos para generar las matrices y vectores que representan ese tipo de análisis. Para crear un mapa de conocimiento a partir del fichero obtenido de WoS hay que realizar los siguientes procesos: preparar los datos exportados desde WoS, extraer el campo a analizar, calcular sus frecuencias, analizar sus co-ocurrencias y preparar la matriz. Después, los resultados obtenidos se dibujan con el programa Pajek, que permite la representación de redes sociales (Bonacich, 2008).

Las instrucciones, absolutamente exactas, para conseguir un mapa de conocimiento se detallan en los siguientes párrafos. Si en las múltiples ventanas de diálogo que emergen en Bibexcel se introducen opciones distintas a las indicadas no se obtendrán los mismos resultados.

Una vez instalado el programa (<http://www8.umu.se/inforsk/Bibexcel/> (<http://www8.umu.se/inforsk/Bibexcel/>)) hay que cargar el archivo con los datos exportados desde WoS y prepararlos para Bibexcel. Según indican Peset y González (2012), tras ejecutar Bibexcel, en la ventana "Select file here" hay que navegar hasta la carpeta correspondiente y seleccionar el archivo "savedrecs.txt" pulsando sobre él dos veces el botón izquierdo del ratón (Figura 1).

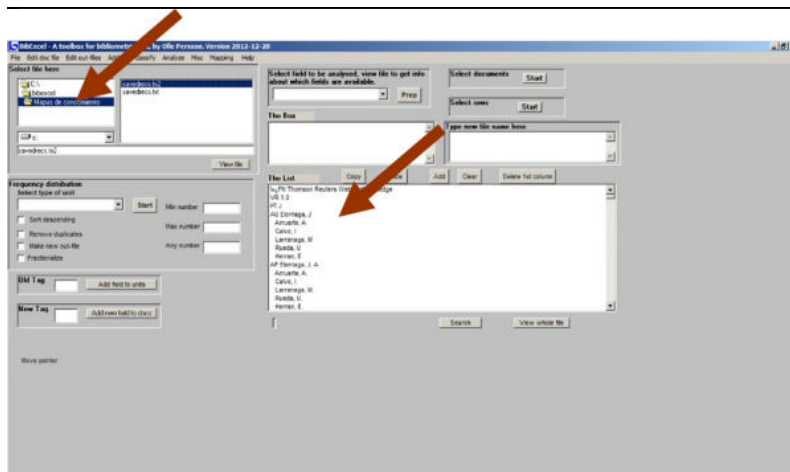


Figura 1. Interfaz de Bibexcel y ejemplo de registro tipo. Las flechas señalan los lugares donde se selecciona el fichero y el lugar donde aparecen los registros

El archivo original de la exportación no será correctamente mostrado en "The list" hasta que se realicen otras dos acciones: introducir los saltos de párrafo e insertar barra finalizadora de campo. Para ello en el menú superior se selecciona "Edit doc-file/Replace line feed with carriage return" y en las ventanas de diálogo que aparecen se debe responder "Yes/Yes" para conservar el txt original y producir un *.tx2.

Para reconocer el formato WoS, se señala el fichero tx2 y se selecciona en el menú superior "Misc/Convert to dialog-format/Convert from WoS" y en la ventana de diálogo se responde "Accept".

De esta acción resulta un nuevo fichero, que será el definitivo, llamado *.doc cuya apariencia es similar al fichero original pero con el retorno de carro en cada campo y un finalizador de línea y de fichero. Se recomienda guardar una copia en lugar seguro, ya que cada vez que Bibexcel realiza una acción se generan nuevas extensiones de fichero, que pueden sustituir al original. Es necesario en este punto reseñar que el documento *.doc no es exactamente un documento de Word. Se trata de un documento en texto plano que el programa le asigna la misma extensión que utiliza Word. Puede ser modificado, con fines de normalización por ejemplo, pero siempre tiene que ser guardado como texto plano.

Seguidamente, para analizar un campo ha de extraerse su contenido. Uno de los datos más adecuados para generar un mapa de conocimiento es el que contiene las ya que al estar normalizadas indicará los temas de forma más precisa que los campos de lenguaje libre, como título o resumen, que requieren un preprocesado más complejo. Para extraer su información se teclea en la ventana "Old tag" la etiqueta que viene asignada por WoS, en este caso DE, y en la ventana "Select field to be analysed" se escoge la opción del desplegable llamada "Any; separated field". Finalmente se presiona el botón "Prep" y en las ventanas de diálogo se procederá con la siguiente secuencia: "Accept/No".

Esta acción genera un fichero *.out² que indica en dos columnas el número de registro donde aparece cada una de las palabras. Estas mismas operaciones hay que realizarlas para extraer cualquier otro campo. Por ejemplo, puede ser de interés el campo título, que también incluye contenido semántico sobre los artículos aunque está redactado en lenguaje natural. Para extraer el campo título, habría que realizar una secuencia similar de acciones escribiendo en este caso la etiqueta TI en el campo Old Tag y seleccionando la opción "Blank-separated words (e.g. title)".

A continuación se calcula el número de veces que aparece un término en el conjunto de documentos seleccionados. Desde la pantalla principal, teniendo seleccionado el fichero *.out, se introducen estos parámetros: ventana "Frequency distribution/Whole string/Start (con sort descending)/Accept".

Esta acción genera un fichero *.cit que muestra en la ventana "The list" dos columnas con los resultados, siendo la primera la frecuencia de aparición y la segunda la palabra cuantificada. Para utilizar los datos hay que copiar y pegar la información en una hoja de cálculo o en el procesador de textos para maquetar la tabla (de ahí el nombre de Bibexcel, por su gran compatibilidad de formatos entre el programa bibliométrico y la hoja de cálculo Excel).

Como hemos mostrado en el anterior apartado, hasta aquí podríamos haberlo hecho con otras herramientas, algo que resulta imposible para el cálculo las co-ocurrencias o relaciones entre los temas. La co-ocurrencia indica el número de veces que dos cadenas de caracteres coinciden en más de un documento. Por tanto, nos permite un análisis sofisticado que establece las relaciones entre términos y las recuenta. Para calcular co-ocurrencias con Bibexcel hay que seleccionar el archivo *.out, escoger en el menú superior "Analyze/Co-occurrence/Make pairs via listbox", y en los cuadros de diálogo presionar "No/Accept". Esta acción genera un fichero *.coc que indica en tres columnas el número de veces que aparecen juntas en diferentes documentos las dos palabras clave.

Este fichero *.coc que contiene las co-ocurrencias necesita de un último paso para poder ser leído en el programa Pajek. Una vez seleccionado el fichero hay que buscar el menú "Mapping/Create net file for Pajek, etc." Esta acción genera un fichero *.net.

Este fichero *.net contiene la matriz numérica de datos que refleja las relaciones establecidas para que puedan ser dibujadas posteriormente. Bibexcel guarda los vértices y su relación (*edges*) en forma de fichero de texto, disponiendo primero la información de los vértices y luego las relaciones entre vértices y su intensidad. En la tabla 1 se puede ver un ejemplo de fichero listo para ser representado. Como se observa, inmediatamente debajo de "edges" aparece la relación entre dos vértices (1 "Data mining"; 2 "New product development", primera y segunda columna de "edges" respectivamente) y número de veces que aparece esta co-ocurrencia (2 veces, tercera columna), formato que Pajek acepta.

* Vértices 820
1 "Data mining"
2 "New product development"
3 "Neural networks"
4 "causal knowledge"
5 "fuzzy cognitive map"
6 "Apriori algorithm"
7 "Clustering analysis"
8 "navigation"
9 "usability"
...
820 "Conceptual frame"
* Edges
1 2 2
1 3 2
4 5 2
6 7 2
6 1 2
8 9 2
10 11 2
...

Tabla 1. Ejemplo de red de co-ocurrencias preparada para su lectura en Pajek

Para visualizar el mapa hay que descargar el software desde <http://pajek.imfm.si/doku.php?id=download> (<http://pajek.imfm.si/doku.php?id=download>) , y abrir la matriz desde la opción *Networks* utilizando el botón del icono de carpeta y buscando el fichero *.net. En la figura 2 se visualiza el aspecto de la pantalla de entrada de este software.

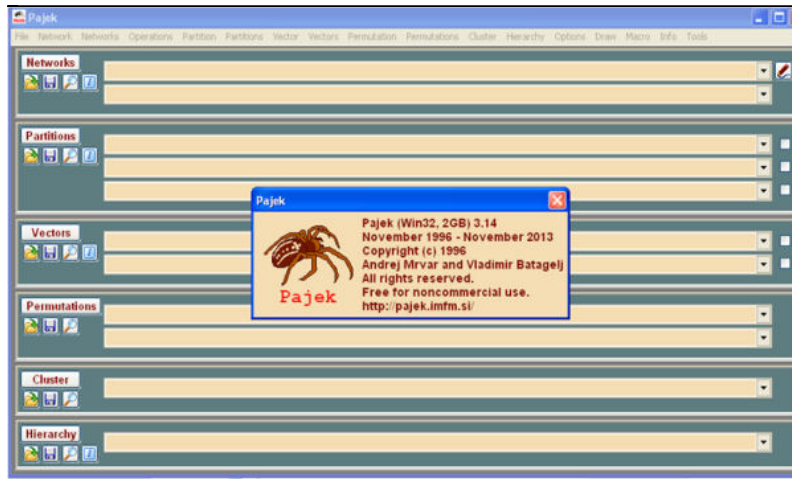
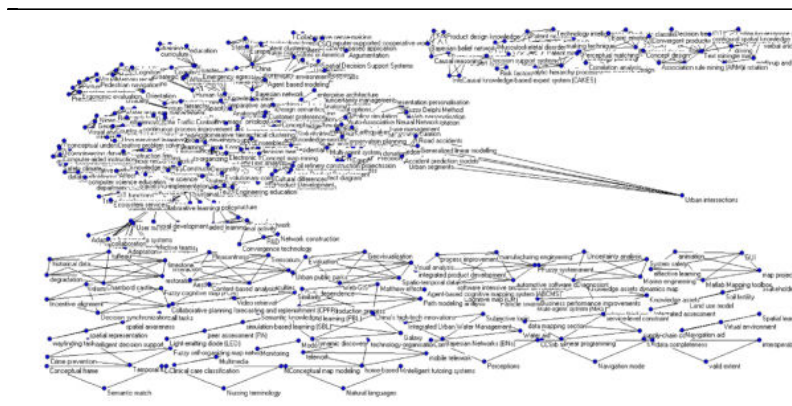


Figura 2. Interface principal de Pajek

Desde el menú superior, la opción "Draw/Network" dibujará el grafo de relación entre los términos en una red circular. En realidad dada la profusión de términos resulta una figura difícil de interpretar, de manera que la separaremos utilizando en el menú "Layout/Energy/Kamada-Kawai/Separated components"(Figura 3).

Figura 3. Red separada con todos los términos de los registros recuperados

Esta visualización permite arrastrar uno de los términos para resaltarlo, pero sigue resultando confusa debido a la cantidad de nodos representados. Para mejorar su interpretación hay que reducir el número de relaciones, tal y como se explica en el siguiente apartado.



4 Mejora de resultados y obtención de los parámetros de centralidad

Una vez vistos los requerimientos mínimos para dibujar el mapa con datos académicos, es posible mejorar su visualización de forma sencilla con las opciones que se explican a continuación.

4.1 Reducir el número de relaciones para interpretar mejor el gráfico

Podemos restringir el número de co-ocurrencias que se visualizan. La ventaja de

hacerlo con Pajek es que se puede ir visualizando la reducción mientras se realiza; sin embargo lo explicaremos en Bibexcel por que el proceso es más sencillo y rápido.

Con el fichero *.coc seleccionado en Bibexcel se añade el umbral mínimo deseado (2 en nuestro caso) en "Min number", que aparece en "Frequency distribution" de la pantalla principal. Después, escogiendo desde el menú superior "Edit out-file/Delete low frequencies" se genera un *.min. Con este archivo se genera de nuevo la red *.net desde el menú "Mapping/Create net-file for Pajek, etc.". indicando en las ventanas de diálogo: "Accept/No/No".

Este último archivo generado se carga de nuevo en Pajek de la manera vista en el apartado anterior. En esta ocasión el resultado es un mapa (Figura 4) más fácilmente interpretable al contar con un número menor de nodos y haberlo separado utilizando en el menú "Layout/Energy/Kamada-Kawai/Separated components".

4.2 Hacer que los nodos sean proporcionales a la frecuencia

Hasta ahora el mapa solo reflejaba las relaciones entre términos. Para visualizar el recuento de aparición de un término se debe crear un vector con las frecuencias. Para ello se selecciona en Bibexcel el *.cit que contenía las frecuencias calculadas de palabras. Se mapea escogiendo en el menú superior "Mapping/Create Vec-file", lo que genera un fichero *.vec.

Se abre en Pajek el fichero *.net desde la opción "Networks" utilizando el icono de carpeta y el fichero *.vec de la misma forma desde la opción "Vectors". Por último, desde el menú superior se utiliza "Draw/Network +First vector" para dibujar la red circular, que podemos separar utilizando en el menú "Layout/Energy/Kamada-Kawai/Separated components".

De esta manera se visualizan desde Pajek los archivos *.net y *.vec a un tiempo en una red con los nodos ponderados por su frecuencia (Figura 4). El tamaño del nodo expresa la frecuencia de aparición, no el número de relaciones. Si escogemos en el menú de la imagen del grafo "Options/Lines/Mark Lines/with Values (Ctrl+V)", se mostrará el número de relaciones sobre las líneas.

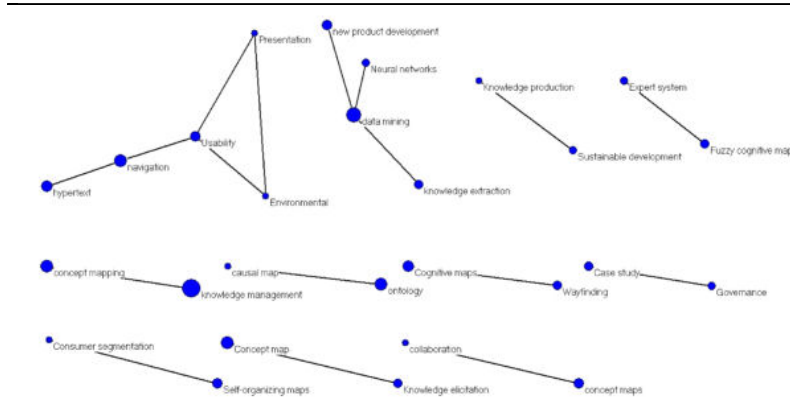


Figura 4. Mapa de co-ocurrencia con un umbral mínimo de 2 y tamaño de los nodos proporcional a su frecuencia de aparición

4.3 Utilizar otro programa de visualización de redes

Vosviewer (versión 1.5.5. Nees Jan van Eck y Ludo Waltman, de la Rijksuniversiteit Leiden) presenta mapas de frecuencias y de co-ocurrencias con otra apariencia debido a que, además de utilizar visualizaciones más sofisticadas, introduce un análisis de probabilidades con el algoritmo matemático llamado Kernel (pestaña "Density view"). La aplicación de este algoritmo, junto con el proceso iterativo que aplica, proporciona no

sólo el recuento sino que, a diferencia de Pajek, también aporta el significado de la proximidad de los términos. Por ejemplo si un término está representado junto a otro, evidencia que ambos están asociados bajo algún criterio de relación, por lo que la distancia en este tipo de mapas es significativa. También representa otros aspectos como agrupación por *clusters* y el tamaño de los nodos según la frecuencia de aparición de los términos.

El programa se descarga gratuitamente desde <http://www.vosviewer.com/download/> (<http://www.vosviewer.com/download/>) y permite cargar redes generadas por Bibexcel o bien crearlas desde el mismo programa, aunque esta última opción trabaja únicamente con los campos de título y resumen.

Esta técnica de minería de texto por tanto es más avanzada que el puro recuento y permite interpretaciones más sofisticadas del mapa, como muestra la pestaña "Density view", que genera una visualización con las tres siguientes variables. En la figura 5 se aprecian las agrupaciones por colores según sea más intensa la frecuencia; en segundo lugar, representa la variación del tamaño de la fuente y de su sombreado; y como última variable presenta la diferente proximidad entre los términos. También puede cargarse la misma presentación por *clusters* calculados a través de un método *k-means*, pestaña "Cluster Density View".

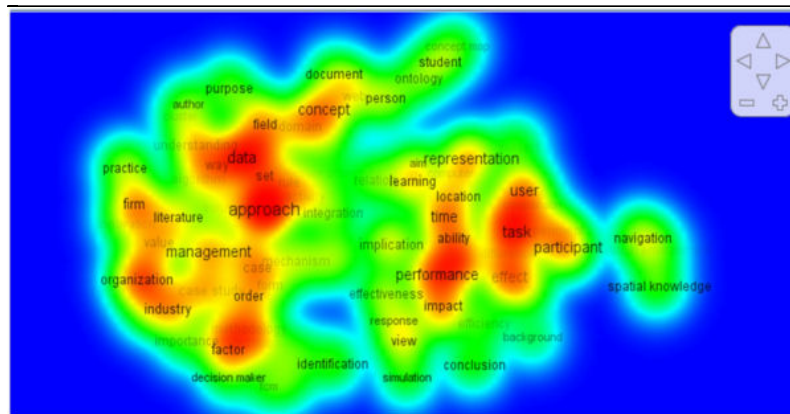


Figura 5. Mapa de densidad generado por Vosviewer

Las instrucciones para conseguir ambas representaciones son muy sencillas e intuitivas. Una vez ejecutada la aplicación escogeremos en el menú de la parte izquierda la opción "Create" y la opción "Create a map based on a network". Tras pulsar en "next" elegiremos la pestaña Pajek e inmediatamente buscaremos el fichero *.net que hemos creado con anterioridad en Bibexcel. A las siguientes preguntas responderemos "Yes/Next/Finish".

A partir de haber generado el mapa, automáticamente el software nos permite visualizar los resultados de 5 formas distintas (*label, density, cluster, cluster density y*

scatter). En definitiva de forma automática Vosviewer ha realizado los cálculos necesarios para generar estas representaciones. Por supuesto el usuario puede guardar tanto los mapas como los datos calculados de forma sencilla en el submenú "Save".

4.4 Cálculo de parámetros de centralidad

Por último queremos hacer hincapié en una cuestión que se suele olvidar: la obtención de parámetros de centralidad. Si bien el presente artículo explica cómo obtener mapas de conocimiento, es necesario reflexionar sobre la utilidad de este tipo de análisis de datos.

A priori la visualización de mapas de conocimiento ofrece al investigador la oportunidad de realizar análisis exploratorios de tipo cualitativo, ya que la interpretación de lo que se visualiza se realiza a través de procesos intuitivos. Podríamos decir que la observación de un mapa de conocimiento aportará una información similar a la contemplación de un cuadro, en la que las conclusiones serán más o menos acertadas en función de la experiencia y pericia del observador.

Para generar un conocimiento más objetivo y desvinculado del grado de conocimiento del investigador, tendremos que realizar cálculos que cuantifiquen de forma objetiva lo que estamos visualizando. Este tipo de análisis está basado en la teoría de grafos, que requiere, al menos, de un conocimiento básico de matemáticas por parte del investigador. No obstante, la interpretación de los parámetros de centralidad de un grafo puede realizarse desde un punto de vista profano.

Son muchos los valores que se pueden calcular en un grafo, pero la tabla 2 presenta algunos de los más usuales en el análisis de redes bibliométricas.

Parámetro de centralidad	Descripción cualitativa	Ecuación
Input degree (grado de entrada)	Número de links que entran en el vértice i .	$d_i^+ = \sum_{j \in G} x_{ji}$
Output degree (grado de salida)	Número de links que salen del vértice i .	$d_i^- = \sum_{j \in G} x_{ij}$
Degree all (grado total)	Número total de links conectados al vértice i .	$D_i = d_i^+ + d_i^-$
Closeness all (proximidad)	Un vértice que se considera importante está relativamente cerrado por otros vértices. Este parámetro indica la cantidad de nodos que hay que recorrer desde la parte externa del grafo hasta llegar al nodo que estamos calculando.	$C_i = \left(\sum_{j \neq i \in G} d(i,j) \right)^{-1}$
Betweenness (intermediación)	Un vértice que se encuentra en vías de comunicación puede controlar el flujo de la comunicación, y por lo tanto es importante. La centralidad por intermediación cuenta el número de caminos más cortos entre i y k que pasan por el actor j .	$B(i) = \sum_{j < k} g_{jk}(i) / g_{jk}$ Donde $g_{jk}(i)$ = número de conexiones geodésicas sobre i , y g_{jk} = número de conexiones geodésicas entre los vértices jk .

Tabla 2. Parámetros de centralidad de un grafo, descripción cualitativa y matemática

Para obtener los valores de centralidad la mejor opción es emplear Pajek, ya que al tener implementadas las ecuaciones descritas en la tabla 2 permite calcularlos de forma sencilla. Un vez el fichero *.net está cargado en Pajek buscamos en el menú superior la opción "Network/Create vector/Centrality" para seleccionar alguna de las medidas de centralidad descritas en la tabla o cualquier otra con la que estemos familiarizados. Los resultados de esta acción se almacenan en forma de un vector de longitud igual al número de nodos que tiene nuestra red. Para consultar los valores de centralidad de cada uno de los nodos, en la parte izquierda de la pantalla utilizaremos el botón "Vectors" para guardar el vector. Este vector puede ser visualizado con cualquier procesador de texto u hoja de cálculo que permite realizar cálculos posteriores.

5 Conclusiones

En definitiva, este texto ha proporcionado los conocimientos prácticos mínimos para comenzar a utilizar programas que han demostrado un gran potencial para el análisis de registros bibliográficos. Básicamente aborda los pasos necesarios para representar la co-ocurrencia de términos, ya que al tratarse de un análisis sofisticado no puede ser dibujado utilizando aplicaciones sencillas. Para finalizar, queremos destacar que estas mismas técnicas pueden ser usadas con textos procedentes de otras fuentes, siempre que se respeten sus requerimientos mínimos: trabajar con texto plano y que los ficheros tengan una estructura comprensible para poder ser computados por Bibexcel. Bajo nuestro punto de vista, cualquier persona que siga estas instrucciones podrá explorar el resto de opciones por sí misma, descubriendo nuevas aplicaciones que van más allá de la bibliometría.

Bibliografía

- Aleixandre Benavent, Rafael; González de Dios, Javier; Alonso Arroyo, Adolfo; Bolaños Pizarro, Máxima; Castelló Cogollos, Lourdes; González Alcaide, Gregorio; Vidal Infer, Antonio; Navarro Molina, Carolina; Coronado Ferrer, Silvia; González Muñoz, M.; Málaga Guerrero, Serafín (2013). "Coautoría y redes de colaboración científica de la pediatría española (2006-2010)". *Anales de Pediatría: Publicación Oficial de la Asociación Española de Pediatría (AEP)*, vol. 78, no. 6, p. 410.
- Alòs-Moner, Adela d' (2003). "Mapas del conocimiento, con nombre y apellido". *El profesional de la información*, vol. 12, no. 4, p. 314-318.
- Ardanuy Baró, Jordi (2014). "Análisis de los estudios bibliométricos en Cataluña". *BID: textos universitaris de biblioteconomia i documentació*, juny, núm. 32 <<http://bid.ub.edu/es/32/ardanuy2.htm> (<http://bid.ub.edu/es/32/ardanuy2.htm>) >. [Consulta: 03/11/2014].
- Bonacich, Phillip (2008). "Exploratory social network analysis with Pajek (http://apps.webofknowledge.com/full_record.do?product=WOS&search_mode=GeneralSearch&qid=4&SID=W2An2nK7qb74jIneZqP&page=1&doc=4)". *Sociological Methods & Research*, vol. 36, no. 4, p. 563-564.
- Cobo, Manuel J.; López Herrera, Antonio G.; Herrera Viedma, Enrique; Herrera, Francisco (2011). "Science mapping software tools: Review, analysis, and cooperative study among tools". *Journal of the American Society for Information Science and Technology*, vol. 62, no. 7, p. 1382-1402.
- Cruz Mata (<http://dialnet.unirioja.es/servlet/autor?codigo=2278510>) , Fermín; Troyano Jiménez (<http://dialnet.unirioja.es/servlet/autor?codigo=860401>) , José Antonio; Enríquez de Salamanca Ros (<http://dialnet.unirioja.es/servlet/autor?codigo=2278508>) , Fernando; Ortega (<http://dialnet.unirioja.es/servlet/autor?codigo=2278512>) , F. Javier (2006). "TextRank como motor de aprendizaje en tareas de etiquetado (<http://dialnet.unirioja.es/servlet/articulo?codigo=2675325>)". *Procesamiento del lenguaje natural* (<http://dialnet.unirioja.es/servlet/revista?codigo=2050>) , núm. 37, p. 33-42. <<http://dialnet.unirioja.es/servlet/ejemplar?codigo=196647> (<http://dialnet.unirioja.es/servlet/ejemplar?codigo=196647>) >. [Consulta: 03/11/2014].
- Del-Fresno-García, Miguel (2014). "Haciendo visible lo invisible: visualización de la estructura de las relaciones en red en Twitter por medio del análisis de redes sociales". *El profesional de la información*, mayo-junio, vol. 23, no. 3, p. 246-252. <<http://dx.doi.org/10.3145/epi.2014.may.04> (<http://dx.doi.org/10.3145/epi.2014.may.04>) >. [Consulta: 03/11/2014].

- Filippov, Sergey (2014). "Mapping Text and Data Mining In academic and Research Communities in Europe". *Lisbon Council special briefing*, no. 16. <<http://www.lisboncouncil.net/publication/publication/109-mapping-text-and-data-mining-in-academic-and-research-communities-in-europe.html> (<http://www.lisboncouncil.net/publication/publication/109-mapping-text-and-data-mining-in-academic-and-research-communities-in-europe.html>) >. [Consulta: 03/11/2014].
- González, Néstor (2013). *Mapas de conocimiento*. <<http://www.innoemotion.com/2013/06/mapas-de-conocimiento/> (<http://www.innoemotion.com/2013/06/mapas-de-conocimiento/>) >. [Consulta: 03/11/2014].
- Hernández Quintana, Ania R. (2006). "Principios ergonómicos aplicados a los mapas de conocimiento: ventajas y desventajas de las nuevas formas de representación de la información". *Acimed*, vol. 14, no. 3. <http://bvs.sld.cu/revistas/aci/vol14_3_06/aci07306.htm (http://bvs.sld.cu/revistas/aci/vol14_3_06/aci07306.htm) >. [Consulta: 19/11/2014].
- Herrero Solana, Víctor; Moya Anegón, Félix (2009). "Redes de coautoría del Departamento de Biblioteconomía y Documentación de la Universidad de Granada (1982-2006)". En: García Caro, Concepción; Vilchez Pardo, Josefina (coord.). *Homenaje a Isabel de Torres Ramírez: estudios de documentación dedicados a su memoria*, Granada: Universidad de Granada, p. 323-332.
- Howard, Ronald A. (1989). "Knowledge maps". *Management science*, vol. 35, no. 8, p. 903-922.
- Leydesdorff, Loet; Wagner, Caroline S.; Park, Han-Woo; Adams, Jonathan (2013). "International collaboration in science: the global map and the network". *El profesional de la información*, January-February, vol. 22, no. 1, p. 87-94.
- López Borrull, Alexandre; Canals, Agustí (2013). "La colaboración científica en el marco de nuevas propuestas científicas: Open Science, e-Science y Big Data". En: Agulló Calatayud, Víctor; González Alcaide, Gregorio; Gómez Ferri, Javier (coord.). *La colaboración científica: una aproximación multidisciplinar*. Valencia: Nau Llibres, p. 91-100.
- Nualart-Vilaplana, Jaume; Pérez-Montoro, Mario; Whitelaw, Mitchell (2014). "How we draw texts: a review of approaches to text visualization and exploration". *El profesional de la información*, mayo-junio, vol. 23, no. 3, p. 221-235. <<http://dx.doi.org/10.3145/epi.2014.may.02> (<http://dx.doi.org/10.3145/epi.2014.may.02>) >. [Consulta: 03/11/2014].
- Peset, F.; Ferrer-Sapena, A.; Villamón, M.; González, L. M.; Toca-Herrera, J. L.; Alexandre-Benavent, R. (2013). "Scientific literature analysis of Judo in Web of Science®". *Archives of budo*, vol. 9, no. 2. <<http://eprints.rclis.org/21008/> (<http://eprints.rclis.org/21008/>) >.
- Peset, Fernanda; González, Luís-Millán (2012). *Construcción de redes de colaboración con Bibexcel, Pajek, Vosviewer y GPSVisualizer*. <<http://personales.upv.es/mpesetm/> (<http://personales.upv.es/mpesetm/>) >. [Consulta: 19/10/2014].
- Pino-Díaz, José; Jiménez-Contreras, Evaristo; Ruíz-Baños, Rosario; Bailón-Moreno, Rafael (2012). "Strategic Knowledge Maps of the Techno-Scientific Network (SK Maps)". *Journal of the American society for information science and technology*, vol. 63, no. 4, p. 796-804.
- Robinson García, Nicolás; **Rodríguez Sánchez** (<http://dialnet.unirioja.es/servlet/autor?codigo=3437540>), Rosa; **García** (<http://dialnet.unirioja.es/servlet/autor?codigo=1800341>), J. A.; Torres Salinas, Daniel; Fernández Valdivia, **Joaquín** (<http://dialnet.unirioja.es/servlet/autor?codigo=585562>) (2013). "Análisis de redes de las universidades españolas de acuerdo a su perfil de publicación en revistas por áreas científicas". *Revista española de documentación científica*, vol. 36, no. 4, 16 p.
- Ruas, Terry Lima; Pereira, Luciana (2014). "Como construir indicadores de Ciência, Tecnologia e Inovação utilizando Web of Science, Derwent World Patent Index, Bibexcel e Pajek? *Perspectivas em Ciência da Informação*, vol. 19, no. 3, p. 52-81.
- Sánchez-Durá, Nicolás (2013). "**Actualidad del relativismo cultural** (http://scholar.google.es/citations?view_op=view_citation&hl=es&user=zftD4skAAAAJ&sortBy=pubdate&citation_for_view=zftD4skAAAAJ:qUcmZB5y_30C)". *Desacatos*, num. 41, enero-abril 2013, p. 29-48
- Tarrats Pons, Elisenda (2012). "Sitkis: una herramienta bibliométrica para el desarrollo del estado de la cuestión". *BiD: textos universitaris de biblioteconomia i documentació*, núm. 28. <<http://bid.ub.edu/28/tarrats2.htm> (<http://bid.ub.edu/28/tarrats2.htm>) >. [Consulta: 19/10/2014].
- Torres Salinas, Daniel; Jiménez Contreras; Evaristo; Robinson García, Nicolás (2014). "Tendencias en mapas de la ciencia: co-uso de información científica como reflejo de los intereses de los investigadores". *El profesional de la información*, vol. 23, no. 3, p. 253-258.
- Tsai, Hsu-Hao (2012). "Global Data Mining: An Empirical Study of Current Trends, Future Forecasts and Technology Diffusions". *Expert Systems with Applications*, vol. 39, no. 9, p. 8172-8181.
- Valenciano Valcárcel, Javier; Devís Devís, José; Villamón, Miguel; Peiró Velert, Carmen (2010). "La colaboración científica en el campo de las Ciencias de la Actividad Física y el Deporte en España" *Revista española de documentación científica*, vol. 33, no. 1, p. 90-105.

Notas

¹ Tema: (Knowledge maps) Refinado por: Categorías de Web of Science: (INFORMATION SCIENCE LIBRARY SCIENCE) Período de tiempo: Todos los años. Índices: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, CCR-EXPANDED, IC.

² Es necesario recordar que bibexcel cada vez que realizamos una acción genera un nuevo documento con una nueva extensión. Es habitual que para finalizar todo el proceso se necesite generar más de 10 ficheros. No obstante todos ellos, con independencia de su extensión, son ficheros de texto plano que se pueden abrir y modificar.

Cita recomendada

García García, Alicia; Pardo Ibáñez, Alberto; Ferrer Sapena, Antonia; Peset Mancebo, María Fernanda; González Moreno, Luis M. (2015). «Herramientas de análisis de datos bibliográficos y construcción de mapas de conocimiento : Bibexcel y Pajek». *BiD: textos universitaris de biblioteconomia i documentació*, juny, núm. 34. <<http://bid.ub.edu/es/34/garcia.htm>> [Consulta: 18-06-2015].

