Facultad de
Comunicación y Documentación

UNIVERSIDAD DE GRANADA

GRADO EN INFORMACIÓN Y DOCUMENTACIÓN

TRABAJO FIN DE GRADO

ANÁLISIS BIBLIOMÉTRICO BÁSICO DE LA EVOLUCIÓN CIENTÍFICA DE LA DEEP WEB: WEB OF SCIENCE 2002-2021

Presentado por:

Dª. Elena Segura Díaz

Tutor:

Prof. Dr. Benjamín Vargas Quesada

Curso académico 2021/2022

D./Dña.: Benjamín Vargas Quesada, tutor/a del trabajo titulado **Análisis bibliométrico básico de la evolución científica de la Deep Web: Web of Science 2002-2021** realizado por el alumno/a **Elena Segura Díaz**, INFORMA que dicho trabajo cumple con los requisitos exigidos por el Reglamento sobre Trabajos Fin del Grado en *Información y Documentación* para su defensa.

Granada, 24 de junio de 2022

Fdo.: VARGAS QUESADA BENJAMIN - 24215184W

Firmado digitalmente por VARGAS QUESADA BENJAMIN - 24215184W
Nombre de reconocimiento (DN): c=ES, serialNumber=IDCES-24215184W, givenName=BENJAMIN, sn=VARGAS QUESADA, cn=VARGAS QUESADA BENJAMIN - 24215184W
Fecha: 2022.06.24 12:32:51 +02'00'

Por la presente dejo constancia de ser el/la autor/a del trabajo titulado **Análisis bibliométrico básico de la evolución científica de la Deep Web: Web of Science 2002-2021** que presento para la materia Trabajo Fin de Grado del Grado en Información y Documentación, tutorizado por el/la profesor/a Benjamín Vargas Quesada durante el curso académico 21- 22.

Asumo la originalidad del trabajo y declaro que no he utilizado fuentes (tablas, textos, imágenes, medios audiovisuales, datos y software) sin citar debidamente, quedando la Facultad de Comunicación y Documentación de la Universidad de Granada exenta de toda obligación al respecto.

Autorizo a la Facultad de Comunicación y Documentación a utilizar este material para ser consultado con fines docentes dado que constituyen ejercicios académicos de uso interno.

24 / 06 / 2022

Fecha Firma

AGRADECIMIENTOS

Me gustaría agradecer a mi familia que siempre me han apoyado y confiado en mí para completar mi formación tanto académicamente como personalmente.

A mi tutor, Benjamín Vargas Quesada, quién me apoyo y ayudo a realizar este trabajo, sin el esto no hubiese sido posible.

A mis profesores y profesoras que me encontré durante este periodo académico, gracias por enseñarme nuevas competencias que antes desconocía.

A mis compañeros y compañeras que he conocido a lo largo de todo el curso, aunque hemos sufrido una pandemia, conseguimos al final poder compartir grandes momentos juntos.

¡GRACIAS POR TODO!

RESUMEN

En este trabajo se realiza un análisis bibliométrico básico de la evolución científica de la Deep Web en la producción científica de la Web of Science. Se obtuvieron un total de 298 documentos en el periodo 2002-2021. Durante este periodo de tiempo el tipo de documento más producido son los documentos de actas de congresos. Se identifico que la fuente más citada es *Lecture Notes in Computer Science*, los autores más representativos desde el inicio de la investigación son "Cho, J", "Ntoulas, A" y "Zerfos, P", el documento fuente más citado en WoS es "*ViDE: A Vision-Based Approach for Deep Web Data Extraction"* escrito por Liu et al., (2010) y se encontraron a través de la técnica de clustering, tres principales líneas de investigación. También, a través de un análisis de la estructura de conocimiento, pudimos detectar tres estructuras de conocimiento: conceptual, intelectual y social. Con la estructura conceptual hemos detectado las palabras clave *data cleaning*, *web data integration* y *schema extraction* como futuros frentes de investigación. Por otro lado, con la estructura intelectual, identificamos las relaciones que tienen los autores a través de la co-citación, obteniendo como resultado dos agrupaciones de investigadores que más se relacionan. En cuanto a la estructura social, obtuvimos como resultado que los países que más colaboran son China, Estados Unidos, Canadá y Brasil, y que institución que más colabora es Jilin University.

Palabras clave: Deep Web, Análisis bibliométrico, Producción científica, Biblioshiny, VOSviewer

ABSTRACT

On this paper a basic bibliometric analysis of the scientific evolution of the Deep Web in the scientific production of the Web of Science is carried out. A total of 298 documents were obtained in the period 2002-2021. During this period of time, the most produced type of document is the documents of congress proceedings. It was identified that the most cited source is *Lecture Notes in Computer Science*, the most representative authors since the beginning of the investigation are "Cho, J", "Ntoulas, A" and "Zerfos, P", the most cited source document in WoS is *"ViDE: A Vision-Based Approach for Deep Web Data Extraction"* written by Liu et al., (2010) and three main lines of research were found through the clustering technique. Also, through a knowledge structure analysis, we were able to detect three knowledge structures: conceptual, intellectual, and social. With the conceptual structure we have detected the keywords *data cleaning*, *web data integration* and *schema extraction* as future research fronts. On the other hand, with the intellectual structure, we identify the relationships that the authors have through co-citation, obtaining as a result two groups of researchers that are most related. Regarding the social structure, we obtained as a result that the countries that collaborate the most are China, the United States, Canada and Brazil, and that the institution that collaborates the most is Jilin University.

Keywords: Deep Web, Bibliometric analysis, Scientific production, Biblioshiny, VOSviewer

TABLA DE CONTENIDOS

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

1.- INTRODUCCIÓN

La Deep Web, conocida como “web profunda”, es todo aquel contenido de Internet que no está indexado por los motores de búsqueda (Ciancaglini et al., 2015).

El origen de la Deep Web comienza en los años 70, con la creación de ARPANET (Monroy-González, 2020, p. 3). Se originó debido a los gobiernos y organizaciones que tenían la necesidad de almacenar sus datos en Internet, pero estos no querían que fuera accesible para todo el mundo (Gallardo-Rosales, 2017).

La primera vez que se dio a conocer el término Deep Web, fue por BrightPlanet, en su libro *“The Deep Web: surfarcing hidden value”* en el año 2001. Se le denomino de esta manera, debido a que muchos aseguraban que se trata de un inmenso iceberg (Gallardo-Rosales, 2017).

En la figura 1, muestra como el iceberg se divide en tres niveles; en el primer nivel está la Surface Web, donde los navegadores convencionales son capaces de indexar los sitios web con acceso público. En segundo lugar, estaría la Deep Web, donde encontraríamos el contenido que no pueden indexar los navegadores, dando lugar a tener que dar una credencial para poder acceder, y, por último, el tercer nivel es la Dark web. Se trata de una red encriptada, en la cual solamente puede ser accesible con ciertas herramientas, dando lugar a tener una navegación totalmente anónima (Cloud Center Andalucía, 2022).

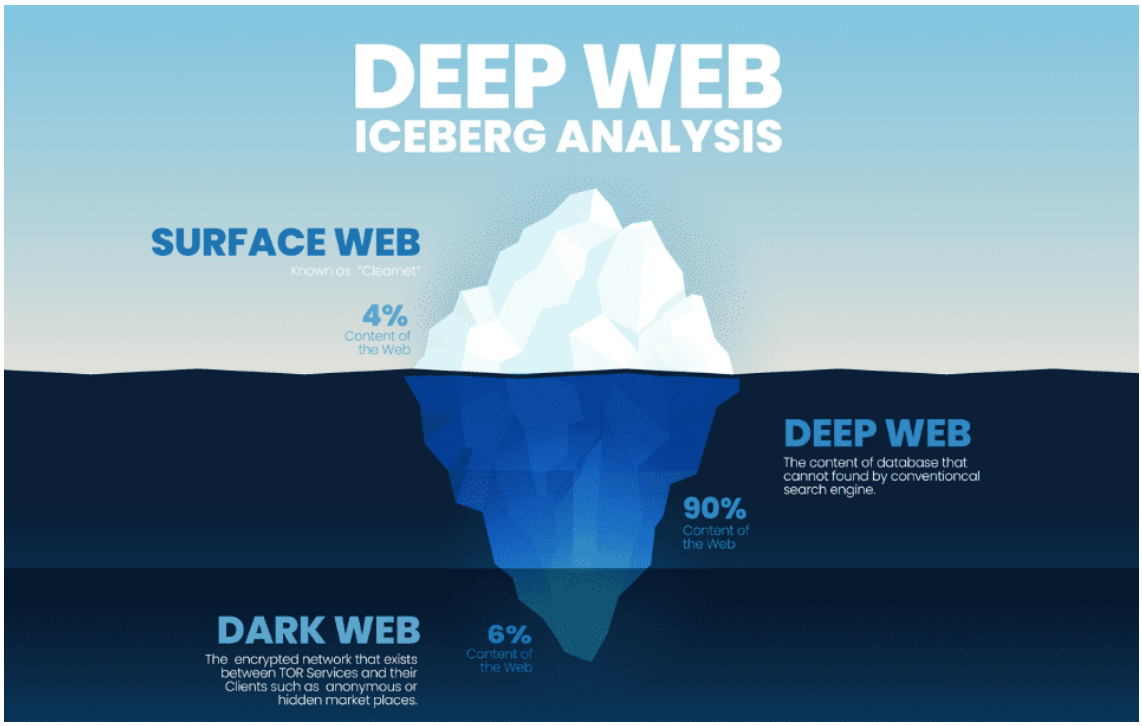


Figura 1. Iceberg Deep Web. Fuente: Cloud Center Andalucía

Aunque el término Deep Web se confunde con Dark Web, ambos términos son distintos. En primer lugar, la Dark Web forma parte de la Deep Web, pero la diferencia que hay es que la Dark Web es mucho más profunda y está basada en una serie de redes anónimas nombradas como darknets (Álvarez, 2018). Estas redes anónimas permiten transferir datos sin que sea detectada la IP (Cloud Center Andalucía, 2022).

Así pues, para poder acceder a la Deep Web, hay varias herramientas que facilitan el acceso, pero en específico la más conocida es The Onion Router (TOR). Es un navegador que te permite acceder a la "Hidden Wiki". La Hidden Wiki es un directorio con varios enlaces de sitios web, los cuales siempre estarán actualizando y cambiando los dominios (Gallardo-Rosales, 2017, p. 3). Además, existe otras herramientas como Freenet y I2P (Invisible Internet Project) (Álvarez, 2018).

Figura 2. Isotipo de The Onion Router, Freenet y I2P

Una vez que tienes acceso a TOR, podemos encontrar todo tipo de información. Por un lado, tenemos información que no se considera ilegal o mala, sino que son de índole privada, secreta o protegida por los gobiernos y organizaciones, también se encuentran foros de diferentes temáticas no accesibles para todo el mundo. Por otro lado, está la parte ilegal, donde encontraríamos por ejemplo, tarjetas de crédito duplicadas, mercado negro, venta de drogas y armas, intercambio de imágenes y archivos censurados, páginas pornográficas de pago y libre acceso, e-books libres con derecho de autor, etc… (Gallardo-Rosales, 2017).

No obstante, esta dificultad de poder acceder a información oculta, ha provocado que los motores de búsqueda se vieran obligados a crear nuevos algoritmos de rastreo para poder indexar todas aquellas redes ocultas que se encuentran en la Deep Web. Esto daría lugar a una cierta investigación, ocasionando una producción de literatura científica (Rai et al., 2020).

En este trabajo hemos intentando realizar un análisis bibliométrico básico para comprobar el crecimiento de la literatura científica de la Deep web. Además, de comprobar qué fuentes, documentos, autores, países y afiliaciones están contribuyendo a la investigación, hemos realizado un análisis de palabras para reconocer qué términos son más utilizados y cuales están surgiendo.

2.- ANTECEDENTES

Hasta donde llega nuestro conocimiento, sólo se han encontrado dos estudios parecidos al nuestro.

El primer estudio corresponde a Rai, S., Singh, K. y Varma, A. K. (2020), quienes realizaron: *"A Bibliometric Analysis of Deep Web Research during 1997-2019"*. El objetivo principal del estudio es realizar un estudio analizando el crecimiento de la literatura de la Deep Web, centrándose en conocer los documentos más citados, identificando autores, afiliaciones, países y palabras clave. Para la realización del estudio, obtuvieron los registros de la base de datos Scopus de los años 1997 hasta 2019, y para analizar los datos obtenidos utilizaron el paquete bibliometrix de RStudio.

El segundo estudio corresponde a Montes Rojano, E. (2019), quien realizó: "*La Deep Web y sus principales líneas de investigación*". Se trata de un trabajo de fin de grado realizado en la Universidad de Granada, concretamente en nuestra facultad. Este estudio pretende identificar cuáles son las principales líneas de investigación de la Deep Web. Primero obtuvieron los registros en la base de datos Web of Science. Una vez extraídos los registros, utilizaron el software CitNet Explorer para identificar cual fue el primer documento que mencionaba la Deep Web. Finalmente, para identificar las principales líneas de investigación, utilizaron el software VOSviewer y CitNet Explorer para el análisis de los registros.

3.- OBJETIVOS

El objetivo general es realizar un análisis bibliométrico básico a partir de la evolución científica de la Deep web en la producción científica de la Web of Science. Este análisis se realizará a través del software Biblioshiny y el software VOSviewer.

Por otra parte, los objetivos específicos son los siguientes. Se dividen en dos grandes grupos:

1. **Análisis de la información**. Donde se identificarán las fuentes, autores, documentos y clustering.
2. **Análisis de la estructura de conocimiento**. Nos encontraremos con la estructura conceptual, social e intelectual.

4.-MATERIALES Y MÉTODOS

4.1. Fuente de datos

Hemos obtenido la información para este trabajo el 30 de marzo de 2022 de la base de datos Web of Science (WoS). WoS pertenece a Clarivate Analytics, y se trata de una base de datos multidisciplinar, donde tenemos acceso a diversas bases de datos y recursos para la investigación (Gutiérrez, 2017). El motivo por el cual fue seleccionada, es debido a su mayor calidad de datos en comparación con Scopus. Aunque sabemos que Scopus tiene

una mayor cobertura de documentos, optamos por tener menos resultados, para lograr una mayor calidad de datos (Aria & Cuccurullo, 2017). La ecuación de búsqueda planteada es la siguiente: *AK=(deepweb) OR AK=(“deep web”) OR KP=(deepweb) OR KP=("deep web"),* obteniendo un total de 298 registros. Aquí tuvimos en cuenta que la ecuación de búsqueda se debía realizar solo en la colección principal (CORE).

Podemos comprobar que hemos utilizado las etiquetas de campo AK y KP. En WoS, podemos buscar por dos tipos de palabras clave. Por un lado, tenemos los Author Keywords (AK). Se tratan de palabras clave que el autor considera que representan mejor el contenido de su artículo (Li, Ding, Feng, Wang, & Ho, 2009, como se citó en Zhang et al., 2016), mientras que los Keywords Plus son palabras clave generadas por un algoritmo informático, el cual se encarga de extraer todas aquellas palabras que aparezcan con cierta frecuencia en los títulos de las referencias de un artículo (Garfield, 1990; Garfield & Sher, 1993, como se citó en Zhang et al., 2016). Incluso, de acuerdo con Garfiled (1990), los Keywords Plus llegan a ser mucho más eficaces a la ahora de representar el contenido de un artículo (citado en Zhang et al., 2016).

Por último, la limitación de años es del 2002 al 2021. Esta limitación comienza desde el año 2002, puesto que es el primer año que WoS tiene un registro que pueda tratar sobre la Deep Web, mientras que el año 2021 es como fecha límite, ya que 2022 aún está incompleto.

4.2. Herramientas

Las herramientas que vamos a utilizar para la realización de nuestro análisis bibliométrico básico, son los siguientes softwares: Biblioshiny (Aria & Cuccurullo, 2017) y VOSviewer (van Eck & Waltman, 2010).

Biblioshiny es una aplicación web que combina las funcionalidades del paquete bibliomatrix de R, con las aplicaciones web que utiliza el paquete Shiny, obteniendo de esta manera un software capaz de realizar el análisis bibliométrico.

Por otra parte, el software VOSviewer, es una herramienta capaz de construir y visualizar mapas bibliométricos con detalle. De esta forma, nos permitirá realizar un mapa de palabras clave de co-ocurrencia y una red de co-citación.

4.2. Datos obtenidos

Los datos que hemos obtenido han sido los siguientes: información principal, evolución de la producción científica anual y el promedio de citas por año.

En la siguiente tabla 1, se identifica la información sobre los datos principales, estando estas divididas en cinco secciones.

Tabla 1. *Información sobre datos principales*

INFORMACIÓN PRINCIPAL SOBRE LOS DATOS	
Periodo de tiempo	2002-2021
Fuentes de información (Revistas, Libros, etc)	247
Documentos	298
Promedio de citas por documentos	3,638
Promedio de citas por año y por documento	0,4074
Referencias	4.555
TIPOS DOCUMENTALES	
Artículo	90
Artículo; capítulo del libro	2
Review	1
Proceedings paper (documento de actas de congresos)	193
Material editorial	2
CONTENIDO DEL DOCUMENTO	
Keywords Plus (ID)	84
Palabras clave de autor (DE)	758
AUTORES	
Autores	622
Apariciones del autor	911
Autores de documentos de un solo autor	29
Autores de documentos de varios autores	593
COLABORACIÓN DE AUTORES	
Documento por autores	0,479
Autores por Documentos	2,09
Co-autores por Documentos	3,06
Índice de colaboración	2,24

En la primera sección, está la información principal sobre los datos, donde podemos verificar algunos datos que son correctos, como por ejemplo el total de documentos descargados fueron efectivamente 298, sin embargo, no sabíamos que 247 son fuentes de información. Por otro lado, tenemos datos relacionados con el promedio de citas que tiene los documentos, siendo un total de 3,638, mientras que el promedio de citas por año y por documento es de un 0,4074. Por último, nos muestran un total de 4.555 referencias, es un dato importante, puesto que son de vital importancia para la elaboración de los documentos y verificación de la información.

En la segunda sección, se identifican los tipos documentales obtenidos. En este caso es curioso que el tipo documental más tratan este tema de investigación sean los documentos de actas (proceedings paper) con 193, mientras que artículos tan solo hay 90. Esto puede deberse a que este tipo investigación se dan más en actas de congresos.

En la tercera sección, está el contenido del documento, este se averigua a partir de las palabras clave de autor y las keywords plus, en este caso se obtuvieron un mayor número de palabras clave del autor, con un total de 758.

En la cuarta sección, se muestran los autores de los 298 documentos. Podemos destacar que tan solo hay 29 autores que han publicado artículos solos, mientras que 593 autores han publicado con más de un autor. Es habitual y común que los autores obtén por colaborar con otros investigadores para la elaboración de artículos, creando de esta manera relaciones con otros investigadores del mismo ámbito de investigación.

Por último, en la quita sección, refleja datos relacionados con la colaboración de los autores. Como se ha mencionado anteriormente, es habitual que los investigadores colaboren entre ellos, en nuestro caso el índice de colaboración es un 2,24.

Por otra parte, en la figura 3 junto con la tabla 2, se muestra cual ha sido la evolución de la producción científica anual. Comprobamos que se trata de una gráfica que comienza desde el primer año que se publicó el primer documento hasta el último año de publicación. Si observamos con detenimiento, podemos identificar que en el 2004 no se publicó ningún documento, mientras que el año de mayor producción científica fue año 2009 con 41 documentos, pero a partir de este año, comenzó a disminuir la producción científica.

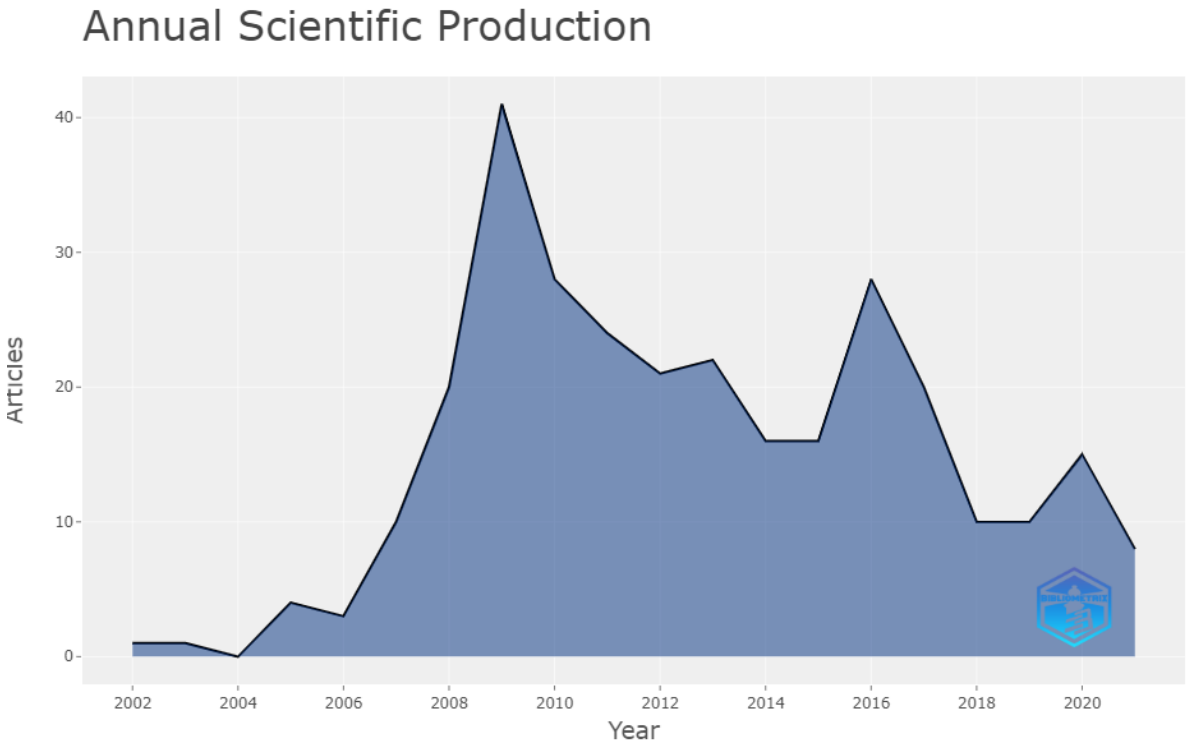


Figura 3. Evolución de la producción científica por años

Tabla 2. *Evolución de la producción científica por años*

Año	Artículos
2002	1
2003	1
2005	4
2006	3
2007	10
2008	20
2009	41
2010	28
2011	24
2012	21
2013	22
2014	16
2015	16
2016	28
2017	20
2018	10
2019	10
2020	15
2021	8

Por último, en la figura 4 junto con la tabla 3, obtuvimos el promedio de citas por año. El primer año con mayor promedio de citas por año fue 2005 y desde entonces ha tenido subidas y bajadas, hasta el año 2021 que se obtuvo un promedio alto. Si observamos la tabla 3, podemos garantizar que en el año 2005 se obtuvo una media de citas por año de un 1,42, mientras que en el año 2021 se obtuvo de media de citas por año un 2.

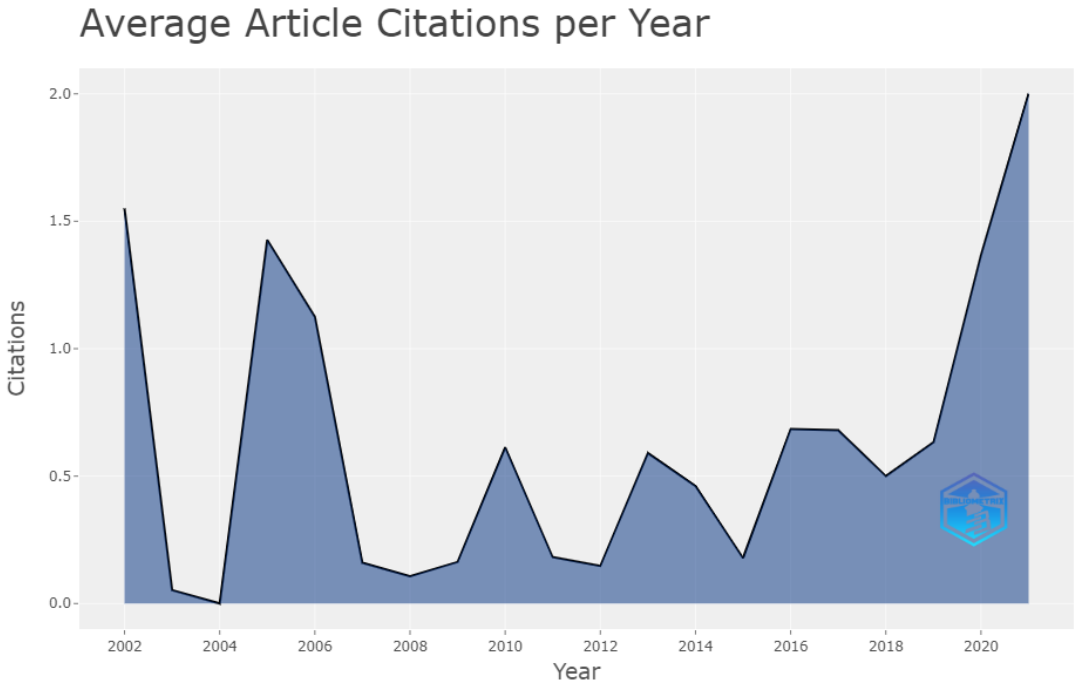


Figura 4. Promedio de citas por año

Tabla 3. *Promedio de citas por año*

Año	Artículos	Media de citas por artículo	Media de citas por año	Años citables
2002	1	31	1,55	20
2003	1	1	0,052631579	19
2004	0	0	0	0
2005	4	24,25	1,426470588	17
2006	3	18	1,125	16
2007	10	2,4	0,16	15
2008	20	1,5	0,107142857	14
2009	41	2,12195122	0,163227017	13
2010	28	7,357142857	0,613095238	12
2011	24	2	0,181818182	11
2012	21	1,476190476	0,147619048	10
2013	22	5,318181818	0,590909091	9
2014	16	3,6875	0,4609375	8
2015	16	1,25	0,178571429	7
2016	28	4,107142857	0,68452381	6
2017	20	3,4	0,68	5
2018	10	2	0,5	4
2019	10	1,9	0,633333333	3
2020	15	2,733333333	1,366666667	2
2021	8	2	2	1

5.- RESULTADOS Y DISCUSIÓN

5.1.- Análisis de la información

A continuación, se van identificar y analizar las fuentes, autores, documentos y clustering.

5.1.1.- Fuentes

Una fuente puede ser una revista, libro, actas de congreso que han publicado uno o varios documentos en la colección bibliográfica (Aria & Cuccurullo, 2017). En los siguientes apartados se van a identificar cuáles son las fuentes de mayor relevancia, las más citadas y las de mayor impacto.

5.1.1.1. Fuentes de mayor relevancia

Las fuentes de mayor relevancia, son aquellas fuentes que han publicado varios documentos. Como podemos observar en la figura 5, hay cinco fuentes de mayor relevancia. Por un lado, tenemos las siguientes revistas: *Data & Knowledge Engineering*, *IEEE Transactions on Knowledge and Data Engineering* y *International Journal of Computer Science and Network Security*. Se tratan de revistas centradas en la ingeniería y conocimiento de datos, informática y seguridad. Y por último, la fuente de mayor relevancia es la *Proceeding of the 22nd International Conference on World Wide Web*.

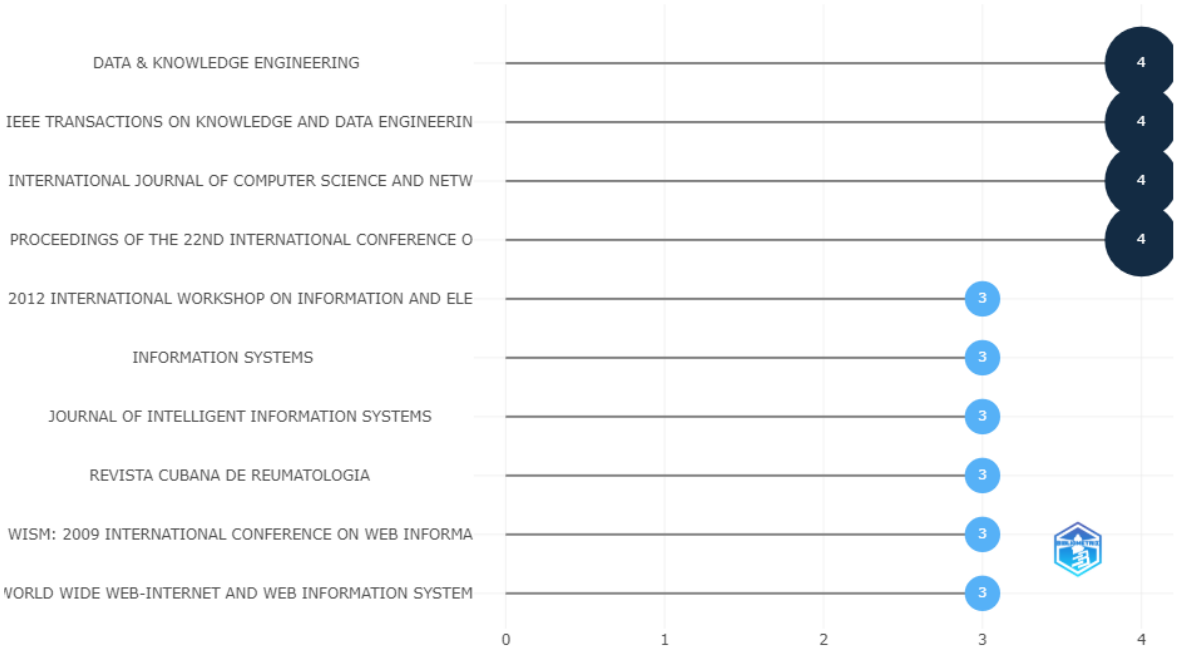


Figura 5. Las fuentes de mayor relevancia

5.1.1.2. Fuentes más citadas en local

Una fuente más citada, es aquella fuente que ha sido citada por uno o más documentos (Aria & Cuccurullo, 2017). En este caso, cuando tratamos con fuentes más citadas en local, esto quiere decir que son las fuentes más citadas de los registros descargados. En la figura 6, se muestran las fuentes más citadas. Por un lado, está *Lecture Notes in Computer Science* con un total de 131 citas, seguida de *Proceedings of the VLDB Endowment* con 122 citas y *SIGMOD Record* con 95 citas.

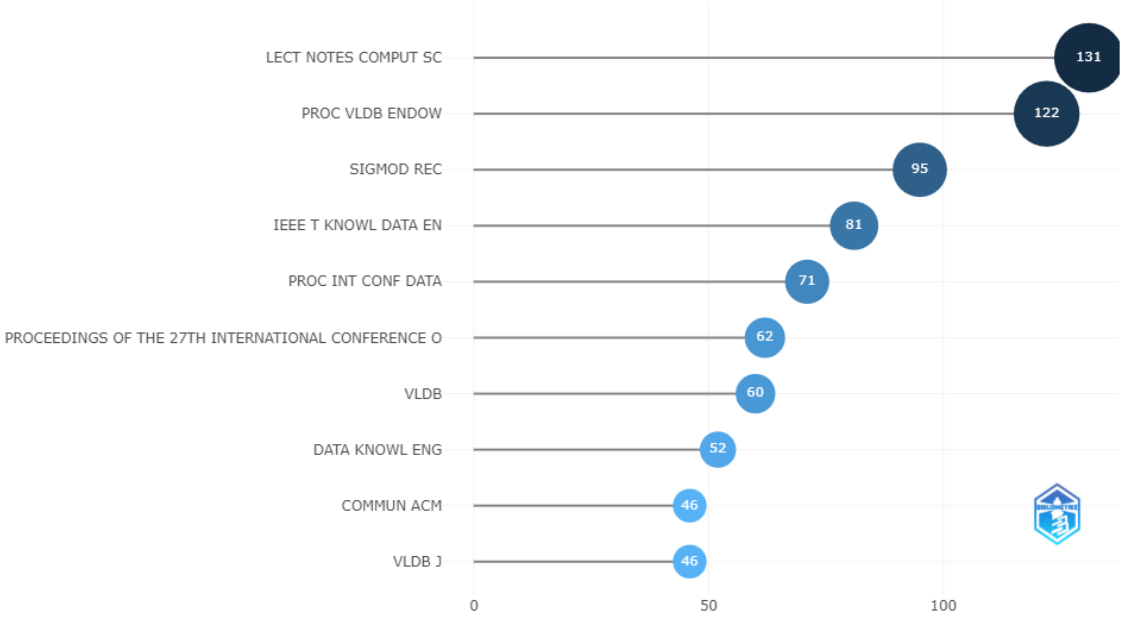


Figura 6. Las fuentes más citadas en local

5.1.1.3. Fuentes de mayor impacto en local

Para comprobar cuales son las fuentes de mayor impacto, hemos utilizado el indicador H-Index. Este indicador lo que pretende es a partir del número autores o revistas, tienen h artículos publicados que han sido citados al menos h veces (Aria & Cuccurullo, 2017). Por lo tanto, como podemos observar en la figura 7, se distinguen cinco fuentes con un mismo impacto. En primer lugar, tenemos las que ya se han reconocido como fuentes de mayor relevancia: *Data & Knowledge Engineering* y *Proceeding of the 22nd International Conference on World Wide Web*. Por otra parte, tenemos la *Information System, 2012 International Workshop on Information and Electronics Engineering* y *World Wide Web-Internet and Web Information Systems.*

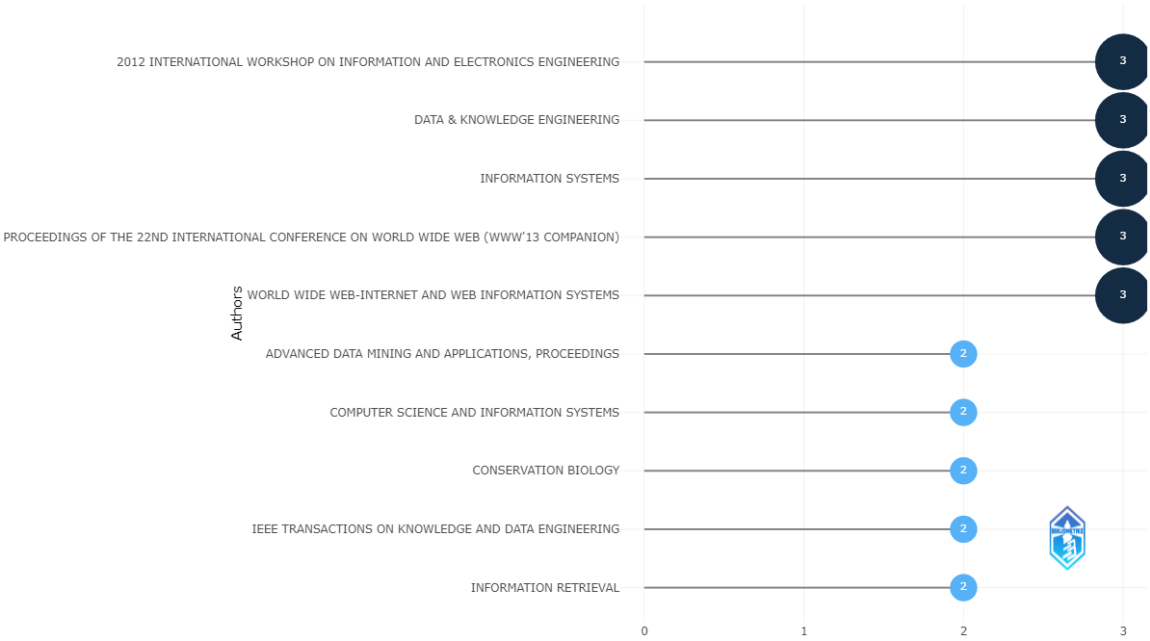


Figura 7. Las fuentes de mayor impacto en local

En la tabla 4, podemos observar que, de las fuentes con mayor impacto, la que más citas ha recibido y más ha publicado es *Proceedings Of The 22nd International Conference On World Wide Web*, sin embargo, la fuente *IEEE Transactions On Knowledge And Data Engineering* aunque tiene un impacto menor, es la fuente que más citas ha recibido.

Tabla 4. *Las fuentes de mayor impacto en local*

Revista	Índice H	Total de Citas	N.º de publicaciones	Año de comienzo de publicación
2012 International Workshop On Information And Electronics Engineering	3	11	3	2012
Data & Knowledge Engineering	3	34	3	2005
Information Systems	3	27	3	2010
Proceedings Of The 22nd International Conference On World Wide Web (WWW'13 Companion)	3	29	4	2013
World Wide Web-Internet And Web Information Systems	3	11	3	2006
Advanced Data Mining And Applications, Proceedings	2	13	2	2008
Computer Science And Information Systems	2	6	2	2011
Conservation Biology	2	55	2	2016
IEEE Transactions On Knowledge And Data Engineering	2	117	4	2005
Information Retrieval	2	28	2	2010

5.1.2.- Autores

Los autores son los creadores de los documentos científicos. Por otro lado, no tan solo se identifican los autores, sino que también las afiliaciones y países de donde se ha publicado el documento. Así pues, en los siguientes apartados se van a identificar cuales son los autores más representativos, los más citados y los que tienen un mayor impacto; y por otro lado se identificaran el país del autor correspondiente, los países con mayor producción científica y los más citados.

5.1.2.1. Los autores más representativos

Los autores más representativos son aquellos autores que tienen un mayor número documentos publicados. En la figura 8, podemos observar, quienes son los 10 autores más representativo. Los autores que más han publicado son “Cui, ZM” y “Wang, Y” con 13 documentos, seguido de “Zuo, WL” con 12 documentos y “Zhao, PP” con 8 documentos.

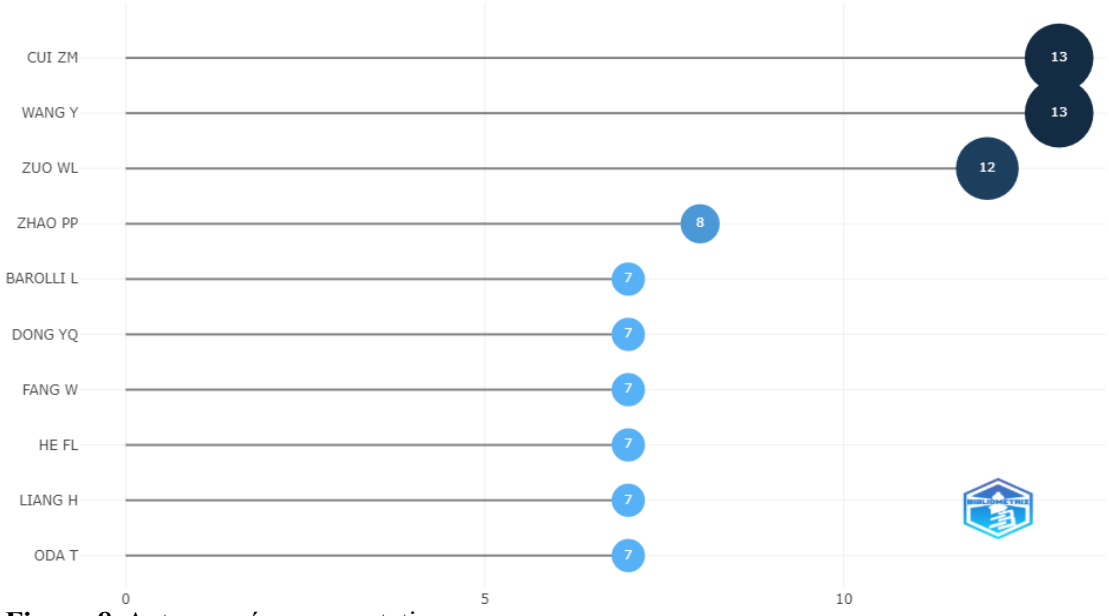


Figura 8. Autores más representativos

5.1.2.2. Los autores más citados en local

Cuando nos referíamos a los autores más citados, son aquellos autores que han recibido un mayor número de citas. Como podemos comprobar, en la figura 9 se muestran cuales son los autores que son más citados: “Cho, J”, “Ntoulas, A” y “Zerfos, P” con un total de 22 citas, posteriormente, los siguientes son “Jiang, L”, “Lui, J”, “Meng, WY”, “Wu, ZH” y “Zheng, QH” con 18 citas.

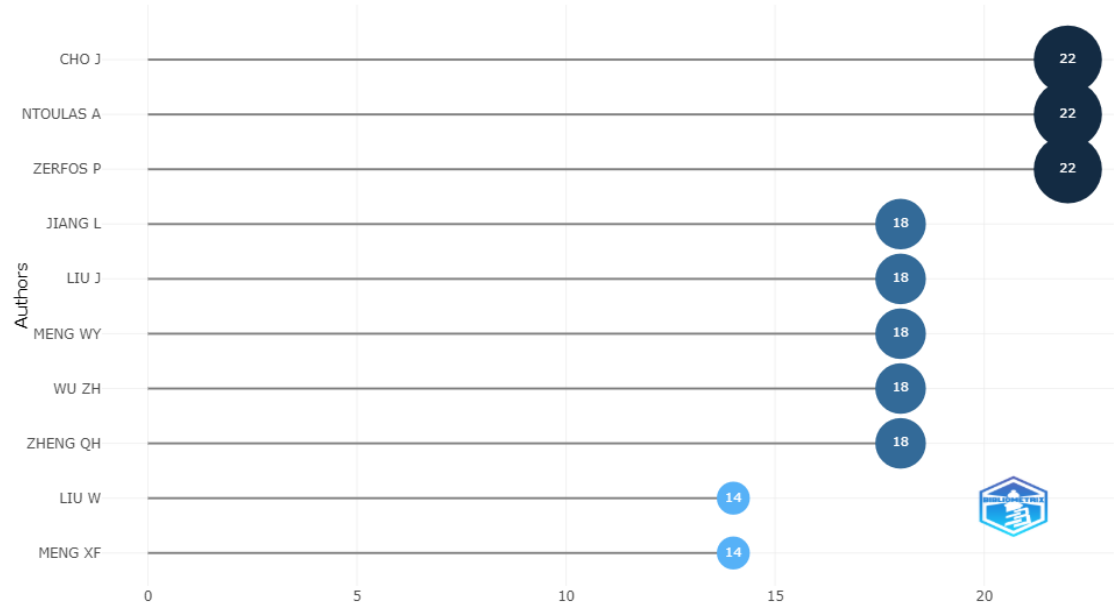


Figura 9. Autores más citados en local

5.1.2.3. Autores con mayor impacto

Para comprobar quienes son los autores de mayor impacto, hemos utilizado de nuevo el indicador H-Index. En la figura 10, se representan los siguientes autores con un mayor impacto. Entre los primeros están los ya mencionados como autores más citados: "Jiang, L", "Lui, J" y "Zheng, QH". Por otro lado, también encontramos con el autor "Wang, Y", ya mencionado como autor más representativo. Y por último, tenemos a los autores "Furche, T", "Lu, JG" y "Wu, ZH". En la tabla 5, identificamos que el autor "Furche, T" además de ser uno de los autores con mayor impacto, se posiciona como el primero debido a su alto número total de citas.

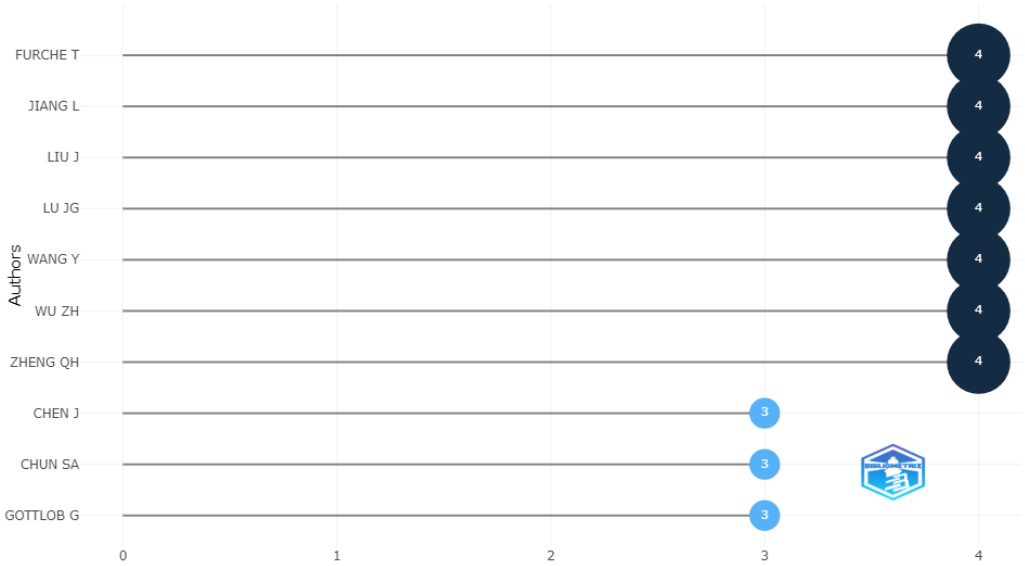


Figura 10. Los autores con mayor impacto

Tabla 5. *Los autores con mayor impacto*

Autor	Índice H	Total de Citas	Nº de Publicaciones	Año de comienzo de publicación
Furche, T	4	66	4	2013
Jiang, L	4	57	5	2009
Liu, J	4	58	6	2009
Lu, JG	4	54	6	2009
Wang, Y	4	44	12	2008
Wu, ZH	4	57	5	2009
Zheng, QH	4	57	5	2009
Chen, J	3	25	5	2009
Chun, SA	3	17	3	2007
Gottlob, G	3	51	3	2013

5.1.2.4. País del autor correspondiente

El autor correspondiente (corresponding author) es el responsable principal de ponerse en contacto con la revista durante el proceso de entrega del manuscrito y publicación del artículo. Los autores son los responsables en proporcionar a la revista los detalles de la autoría y la documentación necesaria para su debida publicación (Elsevier Author Services, 2021), en definitiva son los autores que se atribuyen el mérito de la realización documento. En este caso, se van a comprobar por países que tipo de correspondencia tienen cada uno para identificar que tipo de colaboración tiene cada país.

Si comprobamos en la figura 11, podemos diferenciar dos colores, el color rojo muestra la publicación en varios países (MCP). Este indicador nos muestra la colaboración que tiene con otros países, mientras que el color azul indica la publicación en un solo país (SCP) (Aria & Cuccurullo, 2017). Podemos garantizar que el país que más publica es China y a la vez es el que más colaboración tiene con otros países. Por otro lado, Estados Unidos y India son los que más publican por su cuenta, mientras que Reino Unido es el segundo país que más colaboración tiene con otros países, pero menos documentos ha publicado. Por último, destacar que los únicos países que no han colaborado son India, Alemania y Corea, esto puede deberse a la limitación del idioma y la cercanía.

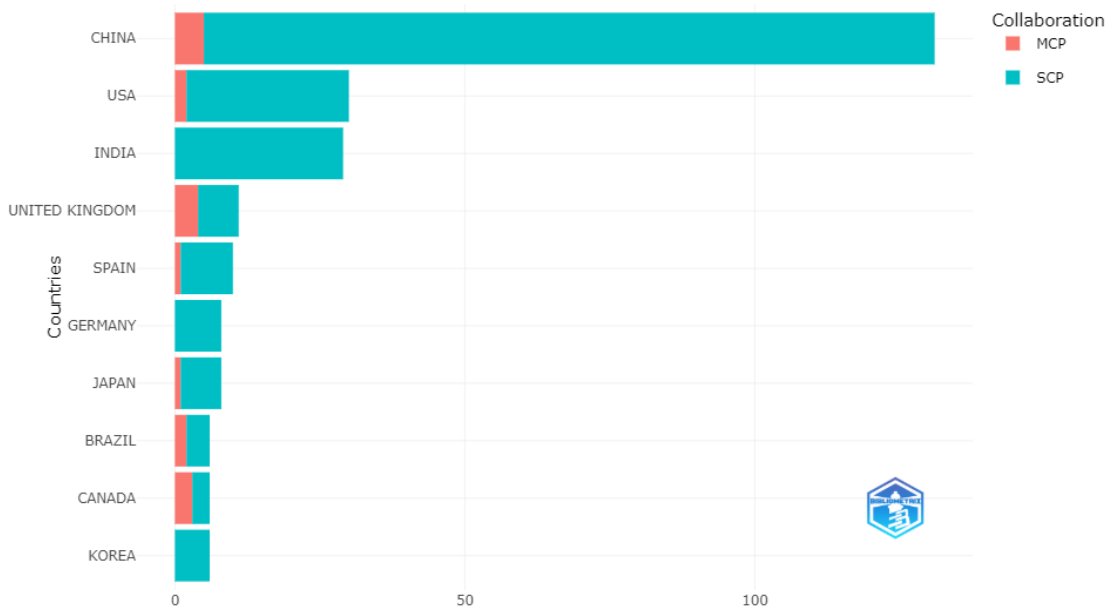


Figura 11. País del autor correspondiente

Tabla 6. *País del autor correspondiente*

País	Artículos	SCP	MCP
China	131	126	5
Usa	30	28	2
India	29	29	0
United Kingdom	11	7	4
Spain	10	9	1
Germany	8	8	0
Japan	8	7	1
Brazil	6	4	2
Canada	6	3	3
Korea	6	6	0

5.1.2.5. Países con mayor producción científica

En la figura 12, se muestra un mapa mundial de los países con mayor producción científica. El mapa se muestra con tonos azules, siendo los tonos más oscuros los países con mayor producción científica. Podemos diferenciar que los países con mayor tonalidad son China y Estados Unidos. En la tabla 7, obtenemos como resultado los 10 países con mayor producción científica, y como se prefijo en la figura 12 los países con mayor producción son China y Estados Unidos.

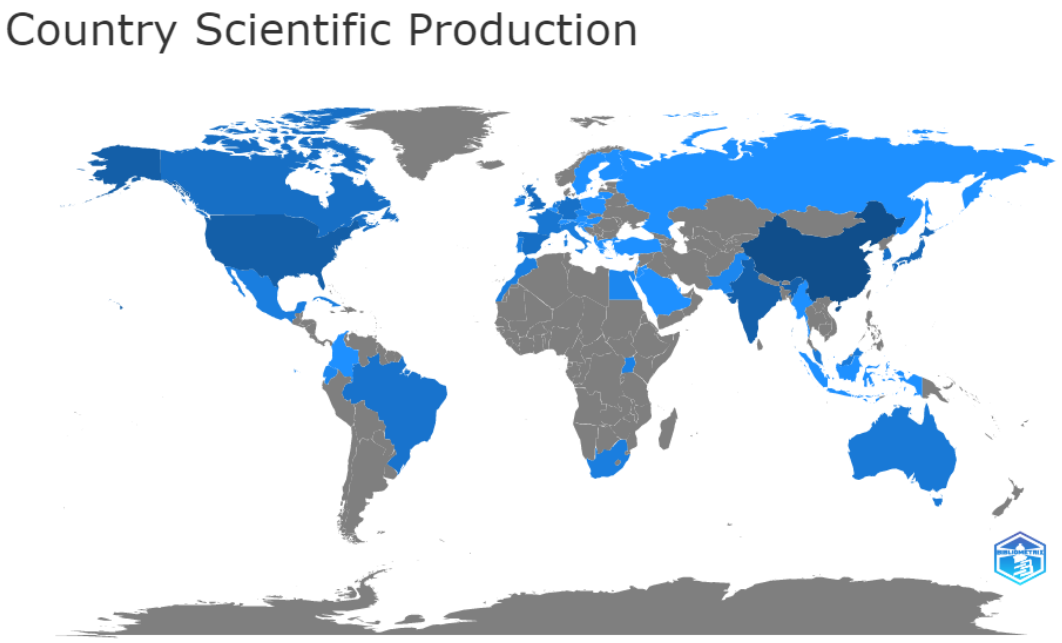


Figura 12. Producción científica por país

Tabla 7. *Producción científica por país*

País	Frecuencia
China	187
Usa	47
India	41
Japan	22
United Kingdom	16
Canada	13
Spain	13
Germany	11
South Korea	11
Brazil	9

5.1.2.6. Países más citados

Los países más citados, son los que han recibido un mayor número de citas. En la figura 13, nos indica que el país que es más citado es China con 354 citas, el segundo es Estados Unidos con 239 citas y el tercero Reino Unido con un total de 239 citas.

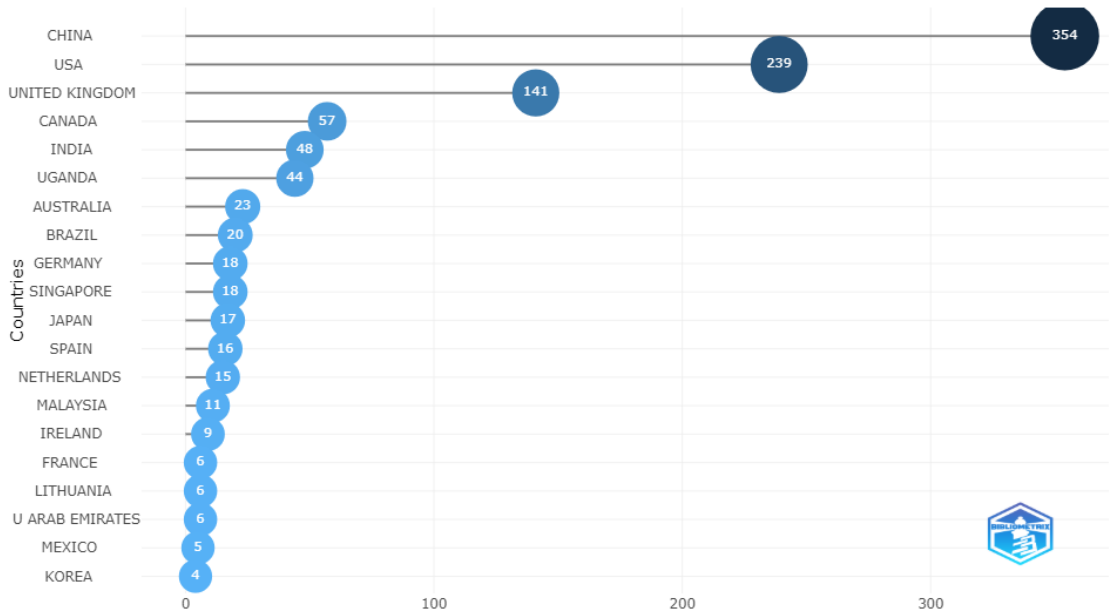


Figura 13. Países más citados.

5.1.3.- Documentos

Los documentos científicos pueden ser por ejemplo un artículo, una revisión o un documento de actas de congresos, y lo que le diferencia de otros es que están incluidos en una colección bibliográfica (Aria & Cuccurullo, 2017). En los siguientes apartados se van a identificar aquellos documentos más citados a nivel global, local y las referencias que más citan.

5.1.3.1. Documentos más citados a nivel global

Un documento citado, es un documento que está incluido en una colección bibliográfica, pero al mismo tiempo es citado en otro documento de la colección. Por lo tanto, cuando nos referíamos a nivel global, es el número de citas que ha recibido un documento de otros documentos incorporados en la base de datos, en nuestro caso en WoS. De esta manera, a través de la base de datos WoS adquirimos los datos del registro de metadatos que previamente se ha descargado. En definitiva, lo que indica es a través de las citas globales, cual es el impacto de un documento dentro de la base de datos bibliográfica (Aria & Cuccurullo, 2017).

En la figura 14, el documento con mayor número de citas a nivel global tiene 104 y se trata del siguiente documento: "*ViDE: A Vision-Based Approach for Deep Web Data Extraction"*, escrito por los autores "Wei Liu", "Xiaofeng Meng" y "Weiyi Meng", y publicado por la revista *IEEE Transactions on Knowledge and Data Engineering* en 2010.

En segundo lugar, el documento más citado tiene 67 citas y es "*Downloading textual hidden web content through keyword queries",* escrito por Alexandros Ntoulas, Petros Zerfos y Junghoo Cho, publicado en *CM/IEEE-CS joint conference on Digital libraries* en 2005.

Por último, el tercer documento más citado tiene 48 citas y es "*Automatic complex schema matching across Web query interfaces: A correlation mining approach"*, escrito por Bin He y Kevin Chen-Chuan Chang, publicado por la revista *ACM Transactions on Database Systems* en 2006.

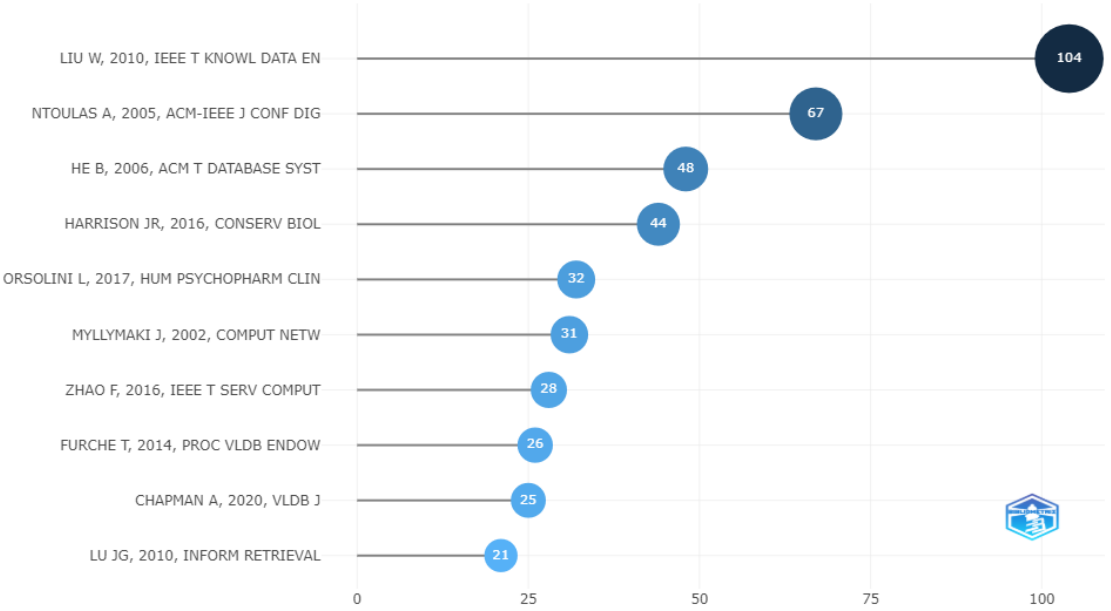


Figura 14. Documentos más citados a nivel global

5.1.3.2. Documentos más citados a nivel local

Para la identificación de los documentos más citados a nivel local, se determina el número de citas que ha recibido un documento a partir de los documentos descargados. Para calcular los documentos más citados, bibliometrix se encarga de analizar todo el conjunto de referencias de los documentos descargados. En resumen, lo que queremos comprobar es cuál es el impacto de un documento dentro de los documentos descargados que se están analizando (Aria & Cuccurullo, 2017).

En la figura 15, el documento más citado a nivel local tiene 22 citas y es el ya mencionado, "*Downloading textual hidden web content through keyword queries",* escrito por Alexandros Ntoulas, Petros Zerfos y Junghoo Cho, publicado en *CM/IEEE-CS joint conference on Digital libraries* en 2005.

En segundo lugar, el documento más citado tiene 14 citas, es también el ya mencionado "*ViDE: A Vision-Based Approach for Deep Web Data Extraction"*, escrito por los autores "Wei Liu", "Xiaofeng Meng" y "Weiyi Meng", y publicado por la revista *IEEE Transactions on Knowledge and Data Engineering* en 2010.

Por último, el tercer documento más citado tiene 9 citas y es "*The ontological key: automatically understanding and integrating forms to access the deep Web"*, escrito por Tim Furche, et al., publicado por la revista *The VLDB Journal* en 2013.

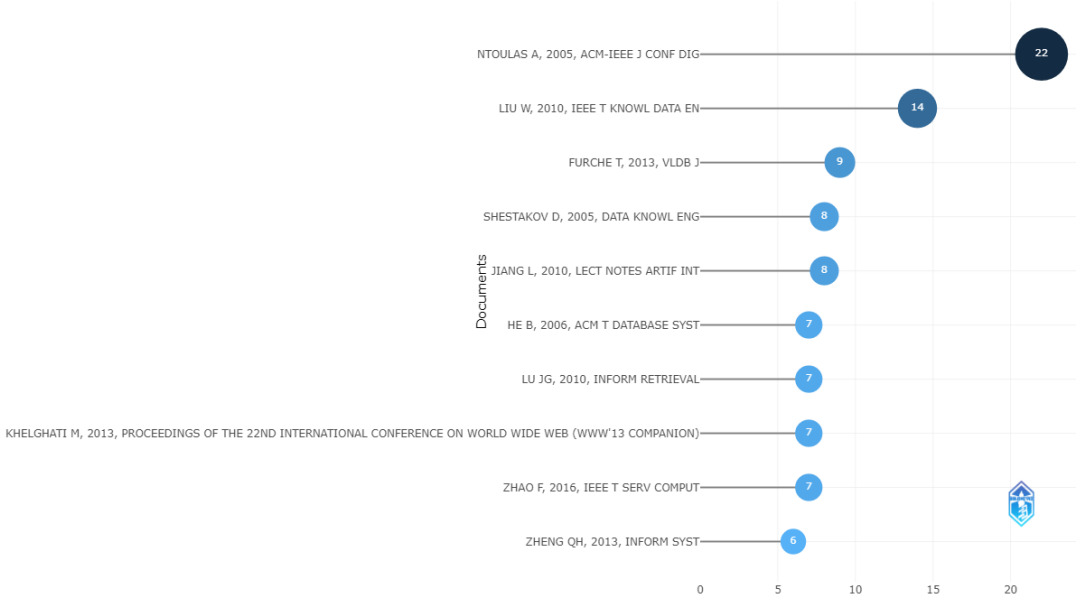


Figura 15. Documentos más citados a nivel local

5.1.3.3. Referencias más citadas en local

Una referencia es cuando un documento científico está incluido en una de las listas de las referencias bibliográficas del conjunto analizado de documentos (Aria & Cuccurullo, 2017). En la figura 16, se indican cuales son las referencias más citadas en local. La referencia que más citas tiene en local, es un documento titulado *"Structured databases on the web: Observations and implications"*, escrito por "Chang, KCC", publicado en la revista *ACM Sigmod Record* en 2004.

Figura 16. Referencias más citadas en local

5.1.4.- Clustering

El clustering es un algoritmo de agrupación capaz de detectar subgrupos de palabras clave que están vinculadas entre sí y que corresponden ya sea a temas de interés o detectar posibles problemas de la investigación. Se pueden realizar diferentes tipos de agrupaciones, pero el más utilizado es el análisis de co-palabras (Cobo et al., 2011).

5.1.4.1. Mapa de clústeres por acoplamiento de documentos

Para la realización del mapa de clústeres por acoplamiento de documentos, primero debemos saber que el acoplamiento (coupling) es cuando dos documentos fuentes citan a un mismo documento. Con la ayuda de Biblioshiny, nos identifica los clústeres de documentos a partir de las palabras clave de los autores junto con la medida de impacto de la puntuación global de citas y etiquetado con las Keywords Plus (Huh, 2021).

Como resultado, en la figura 17 y tabla 8, hemos obtenido un total de siete clústeres de documentos. Podemos observar que los clústeres con mayor cercanía a la centralidad e impacto se encuentran en la parte superior derecha del mapa, siendo estos los clústeres 2 y 3, mientras que los clústeres más alejados de conseguir la centralidad e impacto son los que se encuentran a la izquierda del mapa, los clústeres 7 y 4.

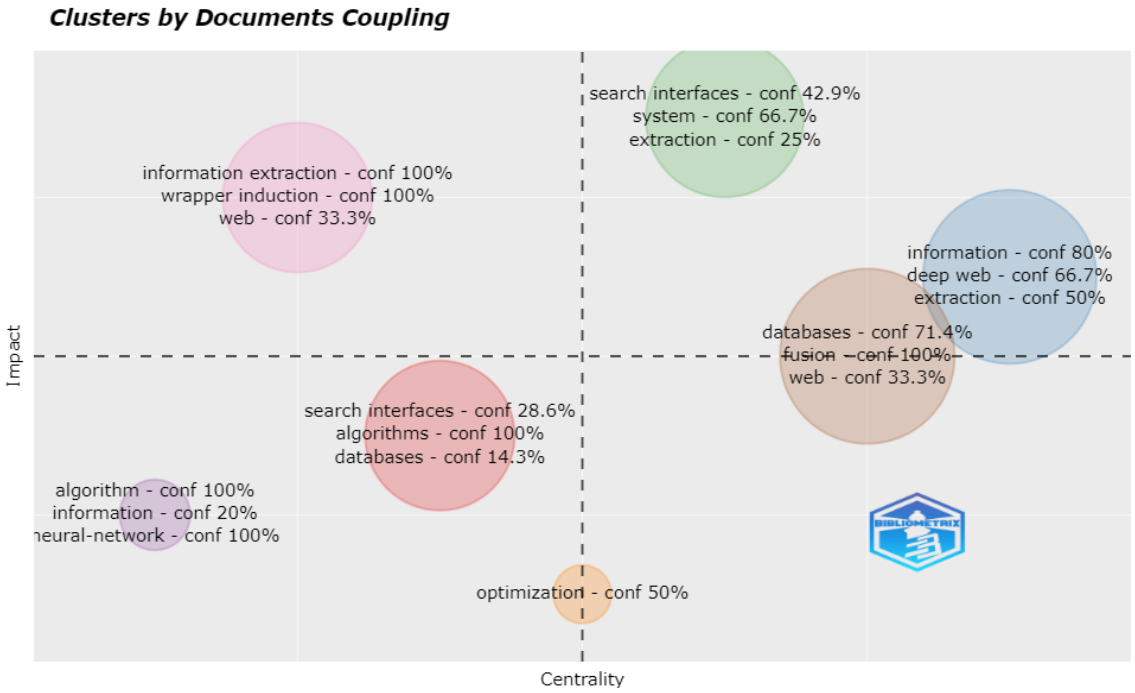


Figura 17. Mapa de clústeres por acoplamiento de documentos

Tabla 8. *Datos del mapa de clústeres por acoplamiento de documentos*

Palabras clave	Clúster	Frecuencia	Centralidad	Impacto
information - conf 80% deep web - conf 66.7% extraction - conf 50%	2	54	0,499184771	1,4635177
databases - conf 71.4% fusion - conf 100% web - conf 33.3%	6	55	0,4427788	1,2616108
search interfaces – conf 42.9% system - conf 66.7% extraction - conf 25%	3	43	0,435979348	1,8861232
optimization - conf 50%	5	10	0,387113012	1
search interfaces - conf 28.6% algorithms - conf 100% databases - conf 14.3%	1	38	0,342686712	1,173807036
information extraction - conf 100% wrapper induction - conf 100% web - conf 33.3%	7	38	0,27545887	1,7526315
algorithm - conf 100% information - conf 20% neural-network - conf 100%	4	12	0,215284511	1,125

5.2.- Análisis de la estructura de conocimiento

Para comprobar cómo ha sido la evolución científica de la Deep web en la producción científica, vamos a realizar lo que se llama un mapeo científico (science mapping). El objetivo que tiene es indicar los aspectos estructurales y dinámicos de la investigación científica (Börner et al. 2003; Morris et al., 2008, como se citó en Aria & Cuccurullo, 2017), pero el mapeo científico se centra principalmente en las estructuras del conocimiento, divididas en tres estructuras: conceptual, intelectual y social (Aria & Cuccurullo, 2017).

5.2.1.- Estructura Conceptual

La estructura conceptual trata de mostrar lo que la ciencia habla según los principales temas y tendencias (Aria & Cuccurullo, 2017). De esta forma, lo que pretende la estructura conceptual es detectar cuales son las relaciones intelectuales entre uno y otros a partir de las palabras clave.

5.2.1.1. Red de co-ocurrencia

Una red de co-ocurrencia es una estructura que nos ayuda a poder comprender los temas cubiertos por un campo de investigación para poder identificar cuales son los temas más relevantes y recientes, esto también se conoce como frente de investigación (Aria & Cuccurullo, 2017).

En la figura 18 se muestra junto con la tabla 9, cuales son las palabras claves más co-ocurrentes. Podemos observar que la palabra clave principal que une con otras palabras es *deep web.* Alrededor de la palabra principal, encontramos palabras clave que están relacionadas con la deep web: *query interface, dark web, invisible web, darknet, surface web, databases, web* y *domain ontology*, mientras que otras palabras clave definen algunas de las funciones y características: *data extraction, information extraction, classification, data integration, ranking, search intrfaces, clustering, intrusión detection* y *deep web data integration*. Todas estas palabras clave son las más utilizadas en la investigación, pero hay otras palabras clave como *data cleaning*, *web data integration* y *schema extraction* que podrán llegar a ser temas emergentes.

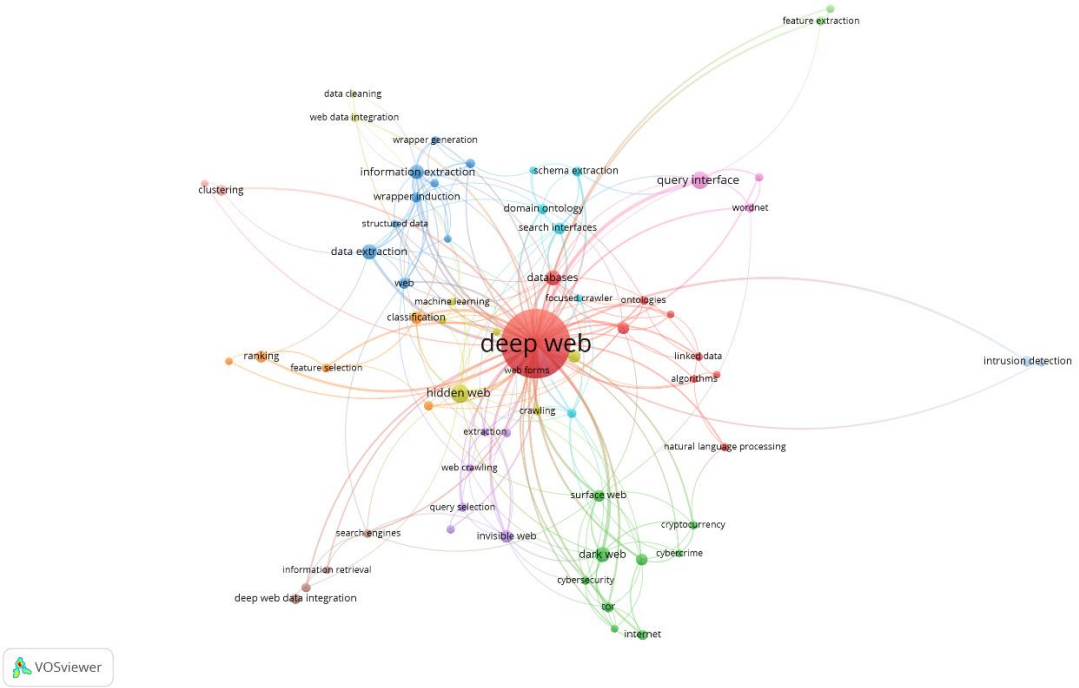


Figura 18. Red de co-ocurrencia

Tabla 9. Palabras clave principales

Palabras clave	Nº de ocurrencias
Deep Web	253
Hidden Web	18
Query Interface	17
Dark Web	12
Data Extraction	12
Databases	12
Information Extraction	11
Classification	9
Data Integration	9
Invisible Web	9
Darknet	8
Ranking	8
Search Interfaces	8
Semantic Web	8
Surface Web	8
Clustering	7
Intrusion Detection	7
Web	7
Deep Web Data Integration	6
Domain Ontology	6

5.2.1.2. Mapa temático

El mapa temático lo que trata de identificar es una red temática sobre una matriz bidimensional, en donde nos muestran dos ejes, el de función de centralidad y densidad investigación (Aria & Cuccurullo, 2017).

La figura 19, es un mapa temático creado para averiguar los temas se tratan en nuestra investigación. El mapa temático se divide en cuatro partes: temas de nicho, temas motores, temas emergentes o de declive y temas básicos. En primer lugar, están los temas de nicho, que son temas que se están tratando, pero no llegan a ser temas básicos o motores, en nuestro caso hay un único cluster que se encuentra como temas de nicho. En segundo lugar, están los temas motores. Se tratan de los temas que más se están investigando, si observamos la figura 17, hay dos clústeres, el clúster de color verde está mucho más elevado, indicándonos que esos son los temas más motores, mientras que el clúster de color morado, está más cerca de que sean temas básicos que temas motores. En tercer lugar, están los temas básicos con un solo clúster. Por último, están los temas emergentes o de declive, en nuestro caso el clúster de color azul, está mucho más cerca de llegar a ser un tema de nicho, mientras que el clúster de color marrón probablemente desaparezca.

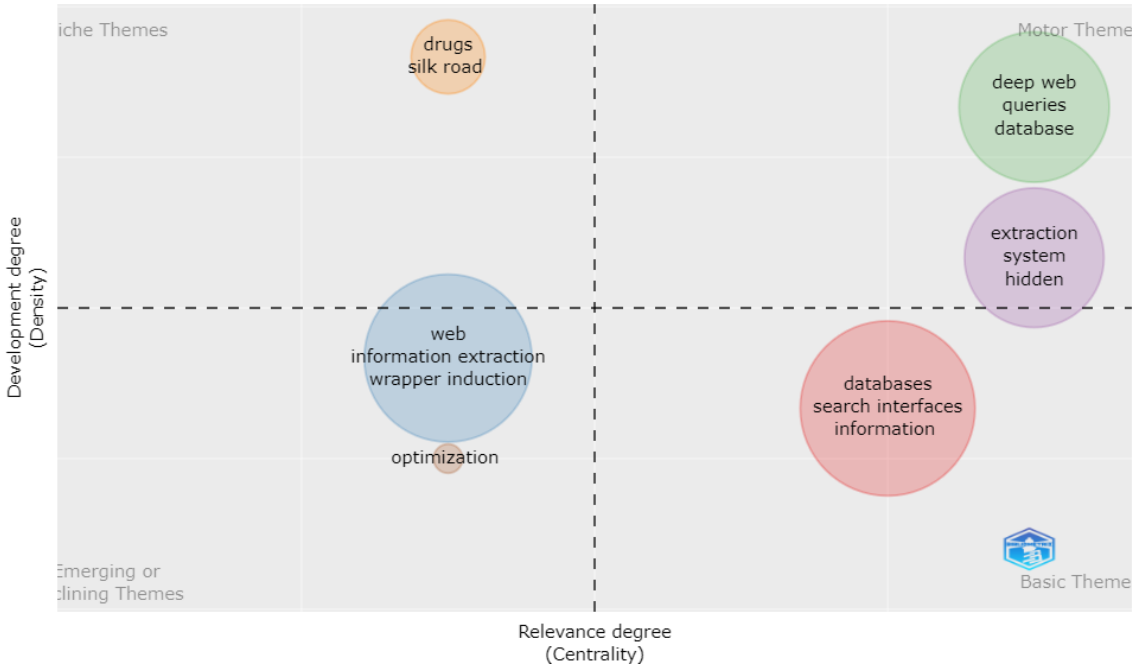


Figura 19. Mapa temático

5.2.1.3. Análisis factorial

El análisis factorial es una técnica estadística de reducción de datos. Esta técnica agrupa las palabras clave por clústeres (Aria & Cuccurullo, 2017). En la figura 20, se ha generado un mapa de la estructurada conceptual, utilizando como método el análisis de correspondencia múltiple (MCA) con las correspondientes Keywords Plus, basándose en el análisis factorial a partir de los documentos extraídos.

Como resultado, nos muestran dos clústeres diferenciados por dos colores, aunque uno de ellos apenas es relevante. Las palabras que se encuentran con mayor proximidad, es porque en los documentos tratan esas mismas palabras juntas, por ejemplo *market, stimulants, impact, psychoactive.substances* e *internet*, representan ciertos documentos que tratan temas relacionados con mercado negro que hay en la deep web. Por otro lado, las palabras clave que se muestran más distantes entre sí, son palabras que muy pocas veces se utilizan (Aria & Cuccurullo, 2017)

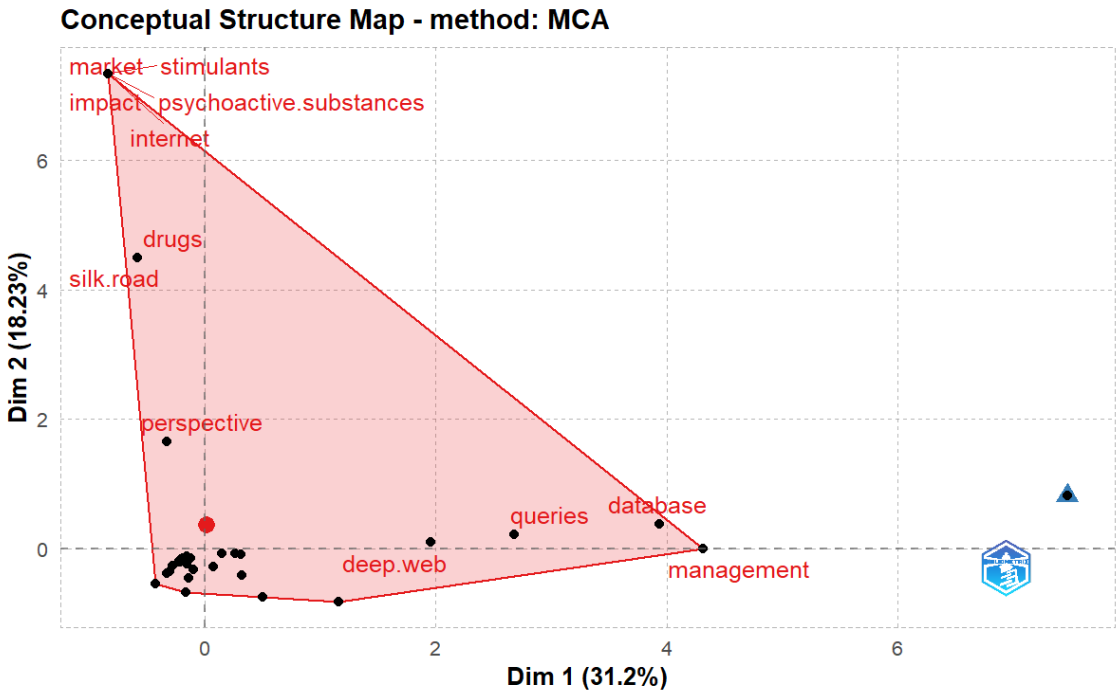


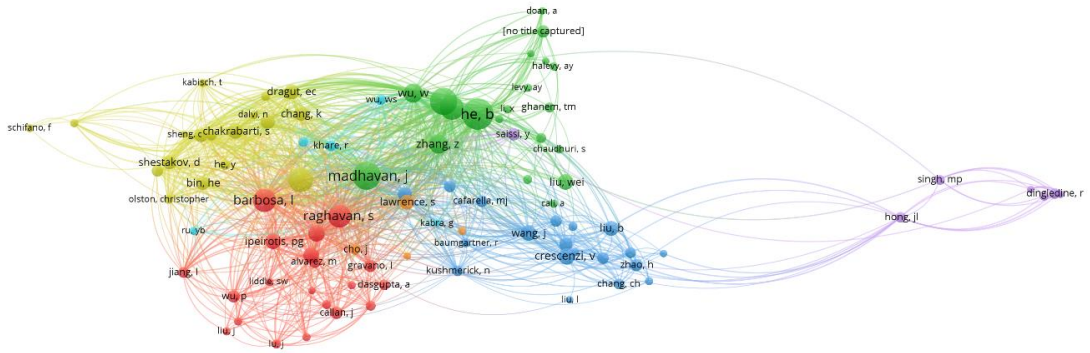
Figura 20. Análisis factorial

5.2.2.- Estructura Intelectual

La estructura intelectual consiste en cómo un documento de un autor influye en una comunidad científica. Para poder comprobar qué relación hay, en bibliometría es muy común identificar las relaciones a través del análisis de las citas, en concreto mediante la co-citación entre autores, documentos y fuentes (Aria & Cuccurullo, 2017).

5.2.2.1. Red de co-citación

La red de co-citación que sea ha realizado, trata de comprobar la co-citación que hay entre los autores. En la figura 21 junto con la tabla 10, se muestran cuales son los autores más co-citados. Los autores con más número de citas pertenecen por una parte al primer clúster mismo clúster: “He, B”, “Madhavan, J” y “Chang, KCC”, y por otra parte al segundo cluster: “Barbosa, L” y “Raghavan, S”.



VOSviewer

Figura 21. Red de co-citación por autores

Tabla 10. Datos de la red de co-citación

Autores	Nº de citas
He, B	124
Madhavan, J	96
Chang, KCC	89
Bergman, MK	77
He, H	69
Barbosa, L	68
Raghavan, S	66
Wu, W	46
Zhang, Z	45
Wang, Y	34
Crescenzi, V	33
Wang, J	32
Ipeirotis, PG	31
Ntoulas, A	31
Chakrabarti, S	30
Liu, Wei	30
Cope, J	29
Dragut, EC	29
Lawrence, S	27
Liu, B	27
Furche, T	25

5.2.2.2. Historiógrafo

El historiógrafo representa cual ha sido el recorrido histórico del tema de investigación, en específico muestra cuales son los autores y documentos principales del recorrido. En la figura 22, observamos que hay una serie de nodos, cada nodo presenta un documento que ha sido citado por otros documentos y cada línea es una cita directa que conecta cada nodo. De esta forma, con el eje horizontal representando los años de publicación, sabremos cual ha sido el recorrido histórico de cada documento y autor (Aria & Cuccurullo, 2017). Los autores que comenzaron el recorrido fueron Shestakov (2005) con el documento *Deque: Querying The Deep Web* y Ntoulas (2005) con el documento *Downloading Textual Hidden Web Content Through Keyword Queries*, siendo este uno de los más citados. A partir del comienzo de estos dos autores, se les fueron sumando más investigadores, pero a lo largo de los años, el autor Liu (2010) con la publicación de *Vide: A Vision-Based Approach For Deep Web Data Extraction*, fue el autor obtuvo más citas globales.

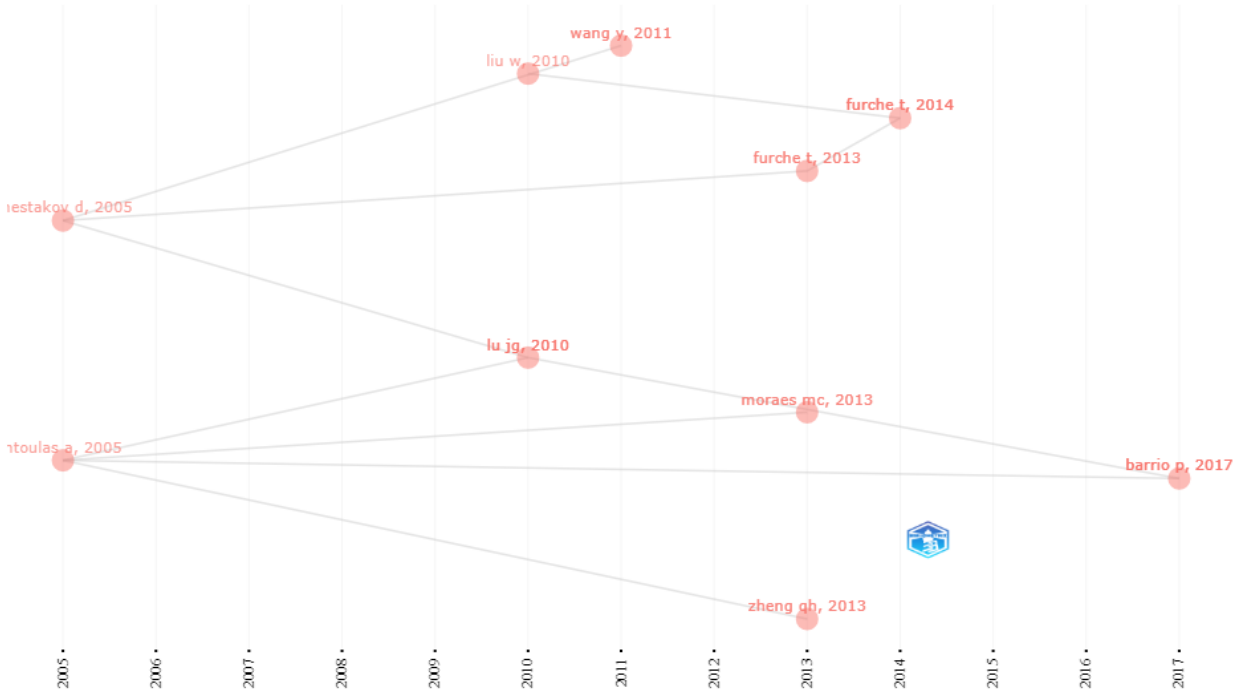


Figura 22. Red histórica de citas directas

Tabla 11. *Datos del historiógrafo*

Autor	Título	Año	Citas locales	Citas Globales
Shestakov, D	Deque: Querying The Deep Web	2005	8	18
Ntoulas, A	Downloading Textual Hidden Web Content Through Keyword Queries	2005	22	67
He, B	Automatic Complex Schema Matching Across Web Query Interfaces: A Correlation Mining Approach	2006	7	48
Lu, JG	Estimating Deep Web Data Source Size By Capture-Recapture Method	2010	7	21
Jiang, L	Efficient Deep Web Crawling Using Reinforcement Learning	2010	8	19
Liu, W	Vide: A Vision-Based Approach For Deep Web Data Extraction	2010	14	104
Furche, T	The Ontological Key: Automatically Understanding And Integrating Forms To Access The Deep Web	2010	9	18
Zheng, QH	Learning To Crawl Deep Web	2013	6	19
Khelghati, M	Deep Web Entity Monitoring	2013	7	7
Barrio, P	Sampling Strategies For Information Extraction Over The Deep Web	2017	3	6

5.2.3.- Estructura Social

La estructura social muestra cómo los autores, instituciones y países se relacionan entre unos y otros en el campo de la investigación científica (Aria & Cuccurullo, 2017).

5.2.3.1. Red de colaboración

La red de colaboración o red de coautoría, es el tipo de estructura social más utilizada (Peters et al., 1991, como se citó en Aria & Cuccurullo, 2017). Este tipo de red nos permitirá comprobar cuales son las relaciones que hay entre instituciones y países.

En la figura 23, se identifican cuales son las instituciones que más colaboran entre sí. Podemos observar se muestran un total de 5 clústeres, de los cuales el clúster con mayor representación es el 1, con la institución Jilin Univ.

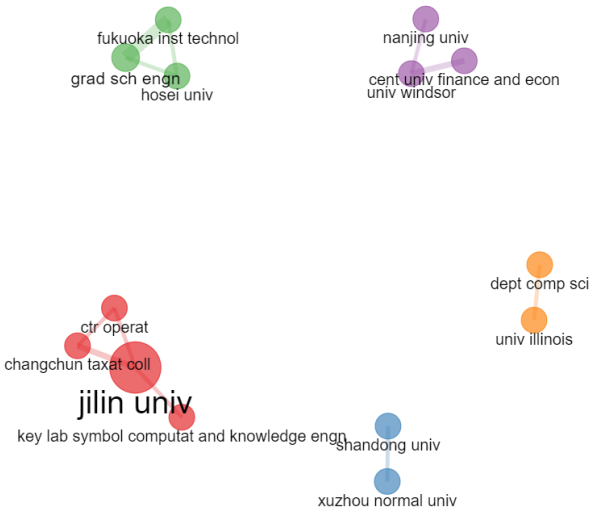


Figura 23. Red de colaboración entre instituciones

En la figura 24, se muestran los países que más colaboran. El país con mayor colaboración es China, junto con Estados Unidos y Canadá, mientras que, por otra parte más europea, se encontraría Italia y Reino Unido.

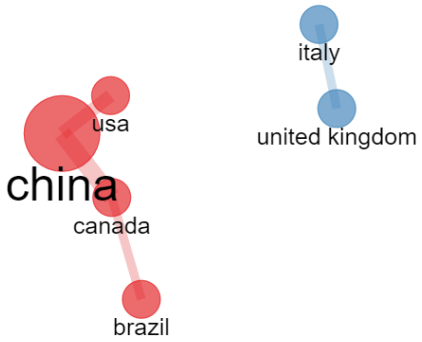


Figura 24. Red de colaboración entre países

5.2.3.2. Mapa mundial de colaboración

La figura 25, muestra un mapa mundial de colaboración, el cual nos resalta con una línea roja la relación que hay entre los países. Podemos deducir que el país que más colabora es China con Estados Unidos y Canadá, pero en la tabla 12, nos muestra que China ha colaborado más con Canadá que con Estados Unidos. Sin embargo, el país que más ha colaborado con otros países es Estados Unidos.

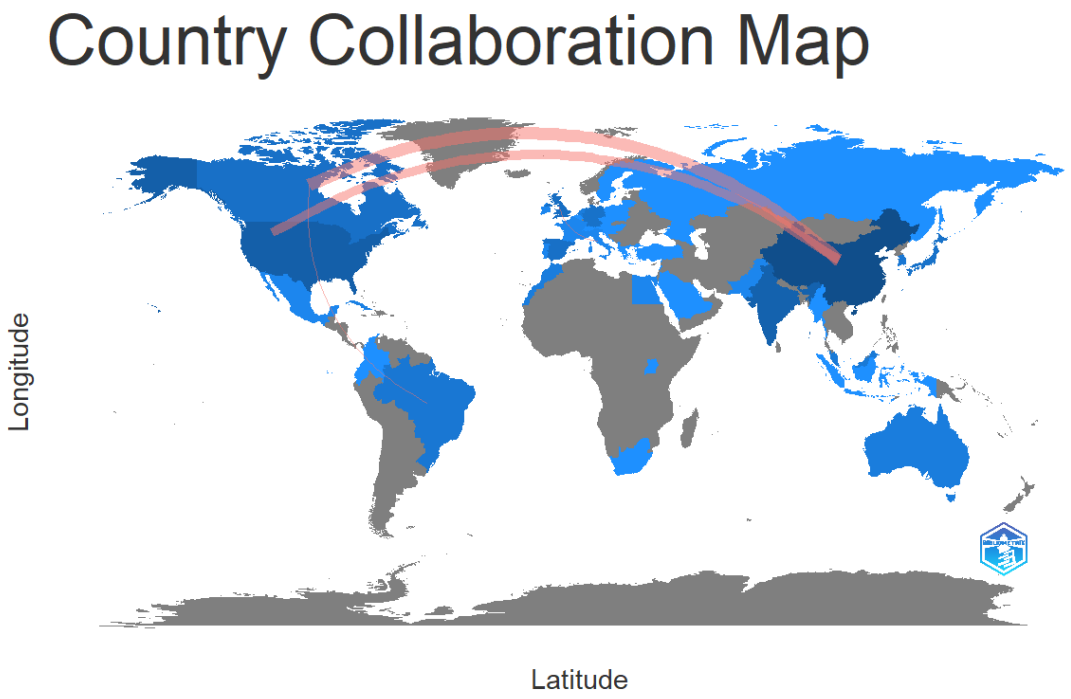


Figura 25. Mapa de colaboración entre países

Tabla 12. Países que más colaboran

Desde	Hacia	Frecuencia
China	Canada	4
China	Usa	3
Canada	Brazil	2
United Kingdom	Italy	2
Canada	Qatar	1
China	Japan	1
Cuba	Ecuador	1
Japan	Albania	1
Jordan	Saudi Arabia	1
Pakistan	Jordan	1

6.- CONCLUSIONES

El tipo de documento más producido en la línea de investigación de *La Deep web,* son los documentos de actas de congresos. Las tres fuentes más citadas, es decir más relevantes, son *Lecture Notes in Computer Science*, *Proceedings of the VLDB Endowment* y *SIGMOD Record*. Los tres autores más citados son "Cho, J", "Ntoulas, A" y "Zerfos, P" y los que más han publicado son "Cui, ZM" y "Wang, Y". El documento fuente más citado por todos los documentos de WoS es "*ViDE: A Vision-Based Approach for Deep Web Data Extraction"* escrito por Liu et al., (2010), mientras que el documento fuente más citado por los documentos descargados en este trabajo es "*Downloading textual hidden web content through keyword queries"* escrito por Ntoulas et al. (2005), ambos documentos tratan de la extracción y descarga de datos en la Deep Web. Sin embargo, la referencia más citada por los documentos fuentes descargados, fue *"Structured databases on the web: Observations and implications"*. Si comparamos estos resultados con el estudio de Rai et al., (2020), en el top 15 de las fuentes prolíficas, las tres más productivas son *Lecture Notes in Computer Science*, *ACM's International Conference Proceedings* y *International Journal of Drug Policy*, en el top 15 de autores altamente productivos, los tres autores más productivos son "Chen, H", "Cui, ZM" y "Bou-Harb, E" y el documento fuente más citado de la base de datos de Scopus es *"Placing Search in Context: The Concept Revisited"* escrito por Finkelstein (2002). En el clustering se identificaron un total de siete clústeres, de los cuales se detectaron las principales líneas de investigación divididas por tres clústeres, en primer lugar, el clúster 2 se detectan las palabras clave *information*, *deep web* y *extracción*, en segundo lugar, el clúster 6, con *databases*, *fusion* y *web*, y en tercer lugar, el cluster 3, *search interfaces*, *system* y *extraction*.

Por otro lado, se hizo un análisis de la estructura de conocimiento para averiguar cuáles eran los aspectos estructurales y dinámicos de la investigación. Para ello, se detectaron tres estructuras: conceptual, intelectual y social. En la estructura conceptual comprobábamos cuales eran los principales temas y tendencias a través de las palabras clave, detectando como tema principal la Deep web, pero centrándose en cuanto a la clasificación, extracción e integración de datos, en herramientas de búsquedas y en bases de datos. En relación con temas que pueden ser tendencia para futuras investigaciones, se detectaron las siguientes palabras clave: *data cleaning*, *web data integration* y *schema extraction.* En la estructura intelectual, identificábamos cuales eran las relaciones que tenían los autores a través del análisis de citas, para lograrlo se realizó una red de co-citación, donde detectábamos dos agrupaciones de investigadores bastante relacionados entre ellos, en la primera agrupación los tres autores que más citas reciben son "He, B", "Madhavan, J" y "Chang, KCC", mientras que en la segunda agrupación los tres autores son "Barbosa, L", "Raghavan, S" y "Ipeirotis, PG". Por último, en la estructura social, comprobamos la colaboración que hay entre países e instituciones. Los países que más colaboran son China, Estados Unidos, Canadá y Brasil, entre estos países el que más ha colaborado es China con Canadá y Estados Unidos. Las tres instituciones que más colaboran son Jilin University, Changchun Taxation College situada en la Jilin University of Finance and Economics y Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education de la Jilin University.

BIBLIOGRAFÍA

Álvarez Rodríguez, J. (2018). *Un paseo por la Deep Web*. [Trabajo Final de Máster, Universidad Abierta de Cataluña]. http://hdl.handle.net/10609/82825%0Ahttp://hdl.handle.net/10609/72810

Aria, M., & Cuccurullo, C. (2017). bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics*.

Ciancaglini, V., Balduzzi, M., Mcardle, R., & Rösler, M. (2015). Below the Surface: Exploring the Deep Web. *Trend Micro.* https://documents.trendmicro.com/assets/wp/wp_below_the_surface.pdf

Cloud Center Andalucía (2022, 27 enero). *Deep Web y Dark Web: qué son y principales diferencias*. Cloud Center Andalucía. Recuperado Abril 22, 2022, de https://www.cloudcenterandalucia.es/blog/deep-web-dark-web-que-son-y-principales-diferencias/

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, *5*(1), 146–166. https://doi.org/10.1016/j.joi.2010.10.002

Elsevier Author Services (2021). *What is a Corresponding Author?*. Elsevier Author Services Blog. Recuperado Mayo 23, 2022, de https://scientific-publishing.webshop.elsevier.com/publication-recognition/what-corresponding-author/

Gallardo-Rosales, R. (2017). La Deep Web. *FIME, Universidad de Colima* https://www.researchgate.net/publication/316884616_La_Deep_Web

Gutiérrez Couto, U. (2017). Guía de uso ISI Web of Science. *Biblioteca Virtual Del Sistema Sanitario Público de Galicia*.

Huh, S. (2021). Document network and conceptual and social structures of clinical endoscopy from 2015 to July 2021 based on the web of science core collection: A bibliometric study. *Clinical Endoscopy*, *54*(5), 641–650. https://doi.org/10.5946/ce.2021.207

Montes Rojano, E. (2019). *La Deep Web y sus principales líneas de investigación.* [Trabajo Final de Grado, Universidad de Granada].

Monroy-González, L. A. (2020). ¿Qué es la Deep Web y qué información podemos encontrar?. *Publicación Semestral*, *3*(5), 1–4. https://repository.uaeh.edu.mx/revistas/index.php/prepa1/issue/archive

Rai, S., Singh, K., & Varma, A. K. (2020). A bibliometric analysis of deep web research during 1997-2019. *DESIDOC Journal of Library and Information Technology*, *40*(2), 452–460. https://doi.org/10.14429/djlit.40.02.15461

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3