

DOI: <https://doi.org/10.56712/latam.v5i4.2458>

## **Análisis bibliométrico de la producción científica de la Deep Web en Ciencias Computacionales**

**Bibliometric analysis of the scientific production of the Deep Web in Computer Science**

**Angel Junior Chacaliza Talla**

achacaliza@undc.edu.pe  
<https://orcid.org/0000-0001-8491-5547>  
Universidad Nacional de Cañete  
Lima – Perú

**Diana Cristina Tello Huapaya**

1970260517@undc.edu.pe  
<https://orcid.org/0000-0001-6411-6959>  
Universidad Nacional de Cañete  
Lima – Perú

**Valeria Elizabeth Arias Pérez**

1971433308@undc.edu.pe  
<https://orcid.org/0000-0001-7152-4934>  
Universidad Nacional de Cañete  
Lima – Perú

Artículo recibido: 18 de julio de 2024. Aceptado para publicación: 01 de agosto de 2024.  
Conflictos de Interés: Ninguno que declarar.

### **Resumen**

El internet está dividido en dos: el espacio virtual, mundo donde nos movemos día a día en la navegación, y el espacio profundo, lugar anónimo que conforma cerca del 96% de todo el internet. La Deep Web ha sido anunciada por muchos como el último bastión de la privacidad en Internet en una era cada vez más intrusiva. Hay muchos estudios sobre las características y aspectos generales de la Deep web, pero ninguno involucró un análisis bibliométrico en el campo de la Ciencias Computacionales. Este trabajo investigó diversas publicaciones en la tecnología de la Deep Web, los datos fueron recopilados desde Scopus y analizados en el software de RStudio. Los resultados muestran a 907 artículos comprendidos entre el periodo de 2008 y 2022. Además, se detalla con claridad que los países con gran cantidad de producción de artículos son: China (37.82%), Estados Unidos (17.42%), India (8.82%) y España (5.07%). Asimismo, se puede observar cómo la productividad anual se ha ido incrementando año tras año teniendo el 2013 la mayor cantidad. Se concluye que, a través de este trabajo de investigación bibliométrica, se descubre el punto crítico de investigación sobre la Deep Web en el campo de Ciencias Computacionales y puede apuntar a futuras investigaciones en el campo. Se espera que más investigadores centren su trabajo en el estudio de la Deep Web, y que esta información se utilice en la creación de un repositorio exclusivo para este tipo de campos poco conocidos en la actualidad.

*Palabras clave:* deep web, internet profundo, análisis bibliométrico, ciencias computacionales

### **Abstract**

The Internet is divided in two: the virtual space, the world where we move from day to day in browsing,

and the deep space, the anonymous place that makes up about 96% of the entire Internet. The Deep Web has been heralded by many as the last bastion of privacy on the Internet in an increasingly intrusive era. There are many studies on the characteristics and general aspects of the Deep Web, but none involved a bibliometric analysis in the field of Computer Science. This work investigated several publications on Deep Web technology, the data were collected from Scopus and analyzed in RStudio software. The results show 907 articles between 2008 and 2022. In addition, it is clearly detailed that the countries with a large amount of article production are: China (37.82%), the United States (17.42%), India (8.82%) and Spain (5.07%). Likewise, it can be observed how the annual productivity has been increasing year after year, with 2013 having the highest amount. It is concluded that, through this bibliometric research work, the critical point of research on the Deep Web in the field of Computer Science is discovered and may point to future research in the field. It is hoped that more researchers will focus their work on the study of the Deep Web, and that this information will be used in the creation of an exclusive repository for this type of currently little known fields.

*Keywords:* deep web, deep internet, bibliometric analysis, computer Science

Todo el contenido de LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades, publicados en este sitio está disponibles bajo Licencia Creative Commons . 

Cómo citar: Chacaliza Talla, A. J., Tello Huapaya, D. C., & Arias Pérez, V. E. (2024). Análisis bibliométrico de la producción científica de la Deep Web en Ciencias Computacionales. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades* 5 (4), 2846 – 2865.  
<https://doi.org/10.56712/latam.v5i4.2458>

## **INTRODUCCIÓN**

Con el nacimiento y evolución del internet en la sociedad se da un origen que al comienzo fue criticado en la historia de la red y sociedad en general, la evolución del internet tuvo grandes impactos dentro de la sociedad, cabe mencionar que “no solo existe una versión del internet, ya que además se pueden encontrar la Surface Web, Deep Web y Dark Web” (Lobo, 2018), (Monroy, 2020). La Deep Web representa un aproximado del 95% de la internet que existe en todo el mundo, la idea que se observa dentro de ella es fomentar el consumo de productos ilícitos, negocios ideales y así como contratar el asesinato a una persona, etc. (Monroy, 2020). La característica principal que posee la Deep Web es el anonimato (Villada & Jiménez, 2017) puesto que se puede navegar por esta web cambiando de IP cada segundo sin logro a localizar el lugar exacto desde donde se navega, ellos el posible con el uso del navegador TOR que tiene como finalidad proteger las comunicaciones de una red repitiéndola en varios servidores al mismo tiempo por todo el mundo, evita que alguien observe los sitios a los cuál se visita y logren localizar la ubicación física. (Courset, Favennec & Hamou, 2017). En la actualidad los pagos que se aceptan en la Deep Web comprenden las criptomonedas, principal medio de pago para los artículos ilegales de la Deep Web, esto se hace principalmente por el anonimato y no dejar huellas. (Monroy, 2020). En la Deep Web se pueden encontrar cosas impactantes tales como drogas, dinero falso, documentación falsa, armas de fuego, órganos humanos, asesinatos en vivo, diferentes tipos de pornografía, entre otros (Pérez, 2018).

A nivel mundial, el campo de las investigaciones a la Deep Web ha ido creciendo considerablemente en los últimos años, como es así en las diversas instituciones y base de datos científicas, y en diferentes idiomas. Es por ello que China se centra en los primeros puestos en realizar investigaciones referidas a la Deep Web en ciencias computacionales, contribuyendo así a realizar diversos estudios de investigación (Toro, 2018).

En Latinoamérica se han comenzado a realizar investigaciones referidas al campo de la Deep Web en ciencias computacionales. Sin embargo, los aportes siguen siendo escasos en las universidades, y aislados por partes de los investigadores, no existe un fin que promueva el intercambio de ideas en estos temas, donde se muestra que solo Estados Unidos con 158 artículos científicos lleva el primer puesto mientras que otros países de Latinoamérica solo cuentan con menos de 5, no existe una región que promueva este tipo de investigaciones (Toro, 2018).

En base a lo mencionado anteriormente, el objetivo principal de la investigación es realizar un análisis bibliométrico de los estudios relacionados a la Deep Web en Ciencias Computacionales, publicados por diferentes autores de diversas IES, pertenecientes a países del nivel mundial, haciendo énfasis en búsquedas en la base de datos Scopus, en el periodo del 2008 hasta noviembre de 2022. Con el fin de conocer y dar nuevos aportes y conocimiento en el tema a nivel mundial, se ha planteado la siguiente pregunta: ¿Cuál es el estado actual de la investigación de la Deep Web en Ciencias Computacionales a nivel mundial?

## **METODOLOGÍA**

Para el desarrollo del trabajo se han realizado una serie de pasos que se detallan a continuación:

### **Alcance de la investigación**

Se seleccionaron dos áreas: 1) el contexto, que consiste en filtrar artículos científicos relacionados con la Deep Web y 2) en el campo de Ciencias Computacionales, que incluye publicaciones de autores relacionados con la IES. A continuación, se construyeron las cadenas con palabras clave, separadas por años de publicación; con el fin de tener un mayor alcance en la búsqueda de información en la base de datos Scopus. La Tabla 1 muestra el alcance de la investigación.

**Tabla 1**

*Alcance de la investigación*

Alcance	Cadena
Contexto Ciencias Computacionales	ALL ("Deep web") AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( PUBYEAR , 2022 ) OR LIMIT-TO ( PUBYEAR , 2021 ) OR LIMIT-TO ( PUBYEAR , 2020 ) OR LIMIT-TO ( PUBYEAR , 2019 ) OR LIMIT-TO ( PUBYEAR , 2018 ) OR LIMIT-TO ( PUBYEAR , 2017 ) OR LIMIT-TO ( PUBYEAR , 2016 ) OR LIMIT-TO ( PUBYEAR , 2015 ) OR LIMIT-TO ( PUBYEAR , 2014 ) OR LIMIT-TO ( PUBYEAR , 2013 ) OR LIMIT-TO ( PUBYEAR , 2012 ) OR LIMIT-TO ( PUBYEAR , 2011 ) OR LIMIT-TO ( PUBYEAR , 2010 ) OR LIMIT-TO ( PUBYEAR , 2009 ) OR LIMIT-TO ( PUBYEAR , 2008 ) ) AND ( LIMIT-TO ( SUBJAREA , "COMP" ) )

### Definición de las preguntas de investigación

Las preguntas de investigación se derivan según el contexto y los países globales, como se muestra en la Tabla 1. La siguiente Tabla 2 define 8 preguntas de investigación que nos permiten definir el estado actual de los trabajos relacionados con la Deep Web en ciencias computacionales.

**Tabla 2**

*Preguntas de investigación sobre Deep Web en Ciencias Computacionales*

Preguntas de investigación	
P1.	¿Cuántos artículos hay sobre Deep Web en Ciencias Computacionales a nivel mundial?
P2.	¿Cuáles son las fuentes más relevantes sobre Deep Web en Ciencias Computacionales a nivel mundial?
P3.	¿En qué idiomas se publican los artículos sobre Deep Web en Ciencias Computacionales a nivel mundial?
P4.	¿A qué países mundiales pertenecen las publicaciones sobre Deep Web en Ciencias Computacionales?
P5.	¿Qué IES a nivel mundial tienen mayor número de publicaciones sobre Deep Web en Ciencias Computacionales?
P6.	¿Cuál es la Producción Científica Anual sobre Deep Web en Ciencias Computacionales?
P7.	¿Cuáles son los artículos más referenciados acerca de la Deep Web en Ciencias Computacionales a nivel mundial?
P8.	¿Qué autores son más relevantes en el campo de investigación sobre Deep web en las Ciencias Computacionales? ?

### Selección de la base de datos

La selección debe proporcionar la documentación más completa posible para lograr responder a las preguntas formuladas en la Tabla 2. Por tanto, en esta investigación se seleccionó la siguiente base de datos:

Scopus: esta fuente de datos fue elegida por las siguientes razones (Semaan, 2018), (Faruk, Rahman, & Hasan, 2021), (Martín-Martín, Thelwall, Orduna-Malea, & Delgado López-Cózar, 2021):

Es la mayor base de datos de investigación de diversos ámbitos, y de alto impacto.

Cuenta con herramientas inteligentes que permiten controlar, analizar y visualizar la investigación académica.

Da todos los metadatos que proporcionan los editores: «autor (es), afiliación (es), título del documento, etc.

Permite limitar los resultados, a través de búsquedas avanzadas (operadores booleanos).

Para extraer los resultados de todos los aportes de la Deep Web, en el ámbito de Ciencias Computacionales, se construyó una cadena de búsqueda que se detalla al final de este párrafo.

En las cadenas de búsqueda se utilizaron las palabras claves en español [Deep Web], y en el ámbito de Ciencias Computacionales; considerando únicamente las publicaciones comprendidas entre el año 2008 hasta el presente año 2022. A continuación, se presenta el criterio de búsqueda elaborado para la consulta en la anteriormente seleccionada base de datos:

```
Scopus: ALL ("Deep web") AND ( LIMIT-TO ( DOCTYPE, "ar" )) AND ( LIMIT-TO ( PUBYEAR, 2022 ) OR LIMIT-TO ( PUBYEAR, 2021 ) OR LIMIT-TO ( PUBYEAR, 2020 ) OR LIMIT-TO ( PUBYEAR, 2019 ) OR LIMIT-TO ( PUBYEAR, 2018 ) OR LIMIT-TO ( PUBYEAR, 2017 ) OR LIMIT-TO ( PUBYEAR, 2016 ) OR LIMIT-TO ( PUBYEAR, 2015 ) OR LIMIT-TO ( PUBYEAR, 2014 ) OR LIMIT-TO ( PUBYEAR, 2013 ) OR LIMIT-TO ( PUBYEAR, 2012 ) OR LIMIT-TO ( PUBYEAR, 2011 ) OR LIMIT-TO ( PUBYEAR, 2010 ) OR LIMIT-TO ( PUBYEAR, 2009 ) OR LIMIT-TO ( PUBYEAR, 2008 )) AND ( LIMIT-TO ( SUBJAREA, "COMP" ))
```

### **Definición de los criterios de inclusión y exclusión**

“La identificación de los criterios es de vital importancia para lograr la aplicabilidad de los resultados de la investigación. Cada uno de los criterios debe de estar definido para que ayuden a identificar los trabajos que cumplen con el objetivo de la investigación” (Manzano & García, 2016).

#### **Criterios de Inclusión**

Permiten filtrar únicamente las referencias de interés que forman parte de los resultados válidos de la investigación (Medina, Marín, & Alfalla, 2010). En este caso, los criterios de selección de los artículos son:

I1: Artículos que analicen el ámbito de la Deep Web

I2: Artículos que sean resultados de investigaciones sobre aprendizaje.

#### **Criterios de Exclusión**

Son las características que no deben tener los documentos, lo que significa que no formarán parte de los artículos seleccionados (Medina, Marín, & Alfalla, 2010). Los criterios de exclusión utilizados para descartar los artículos son:

E1: Artículos que no se encuentren en el ámbito de investigación de Ciencias Computacionales.

E2: Artículos que hayan sido publicados fuera del rango de años de 2008 - 2022.

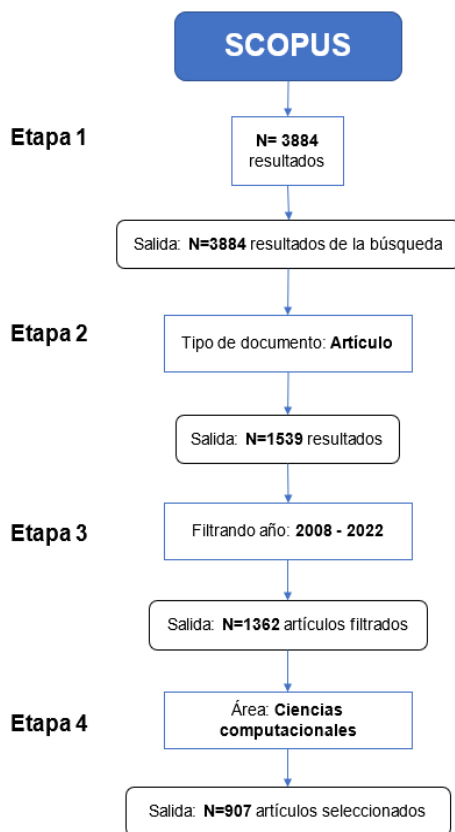
#### **Extracción de los datos**

En la etapa 1, se aplicó la cadena de búsquedas a la Deep Web, donde en una primera instancia se obtuvieron una recopilación de un total de 3884 documentos relacionados al tema anterior (ver figura 1). En la etapa 2, se aplicaron criterios en las cuáles se establecieron que sean de tipo de documento sea solo artículo el cuál se obtuvo una recopilación de 1539 documentos de tipo artículo. En la etapa 3 se filtraron por año comprendidos entre 2008 y 2022 para lo cual se redujo en un 1362 artículos comprendido en aquel rango. En la etapa final se seleccionó que sea del área de ciencias

computacionales. Una vez finalizado el proceso de la etapa de selección, se descartaron un total de 2997 artículos y nuestra muestra se redujo a 907 artículos seleccionados.

**Figura 1**

*Diagrama de selección de art*



## RESULTADOS Y DISCUSIÓN

### Análisis cuantitativo

Respuestas a las preguntas de investigación: Una vez que se dispuso de la lista definitiva de artículos, se seleccionaron los siguientes indicadores: Autor, título, año, tipo de documento, institución, número de citas, país. A continuación se utilizó la herramienta RStudio (library(bibliometrix) - biblioshiny()). Con el fin de poder contestar las preguntas de investigación planteadas en la Tabla 2.

### Análisis general

P1. ¿Cuántos artículos hay sobre Deep Web en Ciencias Computacionales a nivel mundial? En esta investigación se han utilizado 907 artículos de investigación, como se mencionó anteriormente, después de haberle aplicado los filtros necesarios para quedarnos con los artículos más representativos para esta investigación.

### Fuentes de análisis

P2. ¿Cuáles son las fuentes más relevantes sobre Deep Web en Ciencias Computacionales a nivel mundial? A continuación en la tabla 3 podemos ver un listado de los repositorios o fuentes con mayor artículos referentes al tema de la Deep Web; teniendo a JOURNAL OF COMPUTATIONAL INFORMATION SYSTEMS(39) y PROCEEDINGS OF THE ENDOWMENT(33) las fuente con más cantidad de artículos

disponibles, seguidos de IEE ACCESS(20), IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING (18), RUAN JIAN XUE BAQ/JOURNAL OF SOFTWARE (18), JOURNAL OF INFORMATION AND COMPUTATIONAL SCIENCE (17), ACM TRANSACTIONS ON THE WEB (13), INTERNATIONAL JOURNAL OF ADVANCEMENTS IN COMPUTING TECHNOLOGY (13), KNOWLEDGE-BASED SYSTEMS(13) y LECTURE NOTES IN COMPUTER SCIENCE(13).

**Tabla 3**

*Fuentes más relevantes sobre la Deep Web*

Fuente	Número de publicaciones
JOURNAL OF COMPUTATIONAL INFORMATION SYSTEMS	39
PROCEEDINGS OF THE VLDB ENDOWMENT	33
IEE ACCESS	20
IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING	18
RUAN JIAN XUE BAQ/JOURNAL OF SOFTWARE	18
JOURNAL OF INFORMATION AND COMPUTATIONAL SCIENCE	17
ACM TRANSACTIONS ON THE WEB	13
INTERNATIONAL JOURNAL OF ADVANCEMENTS IN COMPUTING TECHNOLOGY	13
KNOWLEDGE-BASED SYSTEMS	13
LECTURE NOTES IN COMPUTER SCIENCE (INCLUDING SUBSERIES LECTURE NOTES IN ARTIFICIAL	13

### **Análisis de idiomas y países**

P3. ¿En qué idiomas se publican los artículos sobre Deep Web en Ciencias Computacionales a nivel mundial? El idioma de la disponibilidad de las publicaciones científicas, se constató que, de los 907 artículos analizados, se encontraron 3 idiomas, destacando el idioma inglés con un total de 839 (92,20%) en la mayoría de las publicaciones. en chino. Hubo 56 (6,15%) publicaciones y sólo 5 (0,55%) en portugués. Se cree que esto se debe a que el idioma inglés es el más utilizado por los investigadores para conseguir una mayor difusión y aceptación de sus trabajos, lo que a su vez supone una mayor posibilidad de publicación en las revistas más prestigiosas.

**Tabla 4**

*Top 10 de países con más publicaciones sobre la Deep Web*

Países	Números de publicaciones
China	343
Estados Unidos	158
India	80
España	46
Reino Unido	41
Alemania	33
Italia	29
Australia	28
Canadá	27
Malasia	24

## Análisis de universidades

P5. ¿Qué IES a nivel mundial tienen mayor número de publicaciones sobre Deep Web en Ciencias Computacionales? En términos de productividad institucional, de las 793 instituciones, la Tabla 5 enumera 15 instituciones con un mayor número de contribuciones a la web profunda en campos de las ciencias de la computación. Mientras que la Universidad Del Noroeste (UNE) de la Ciudad de México aparece en primer lugar con un total de 65 artículos publicados, la Universidad de Jilin (UJ) de China le sigue con 61 artículos, seguida en el mismo orden por la Universidad de Soochow con 58 artículos, con 57 artículos la institución Database Centre for Life Sciences (DBCLS), seguida por la Universidad de Wuhan (WHU) con 50 artículos, también las siguientes instituciones, y finalmente con 18 artículos la Universidad de Oxford en Inglaterra.

**Tabla 5**

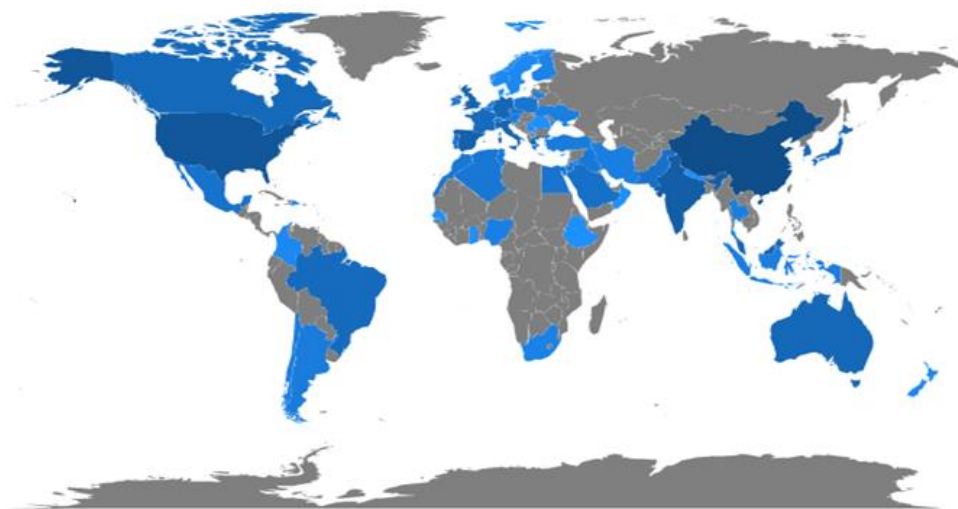
*Universidades con mayor productividad*

Países	Universidades	Números de publicaciones
México	Universidad Del Noroeste (UNE)	65
China	Universidad De Jilin (UJ)	61
China	Universidad De Soochow	58
Japón	Centro De Base De Datos Para Ciencias De La Vida (DBCLS)	57
China	Universidad De Wuhan (WHU)	50
China	Universidad De Shandong	45
China	Instituto De Tecnología De Harbin (HIT)	33
China	Universidad Renmin De China	31
Inglaterra	Universidad De Manchester	27
China	Universidad De Chongqing	23
China	Universidad De Yanshan (YSU)	23
Reino Unido	Escuela De Informática Y Tecnología	22
Estados Unidos	Universidad De Illinois En Chicago	21
China	Universidad De Finanzas Y Economía De Jiangxi (JUFE)	18
Inglaterra	Universidad De Oxford	18



**Figura 2**

*Producción de Artículos por país (color azul: país o región con publicaciones, color gris: país o región sin publicaciones, la intensidad del color: el número de publicaciones)*



**Análisis de publicación anual**

P6. ¿Cuál es la Producción Científica Anual sobre Deep Web en Ciencias Computacionales? La tabla 6, muestra la productividad anual de artículos entre los años 2008 hasta 2022 se ve un crecimiento de producción, en el año 2008 se publicaron un total de 43 (4.7%) artículos, el año 2009 se vio un decrecimiento en el cual solo se publicaron 36 (4.2) artículos, en el 2010 se observa un crecimiento en el cual se publicaron artículos con un total de 71 (7.8%), en el 2011 y 2012 se publicaron artículos con un total de 67 (7.4%) y 62 (6.8%) artículos respectivamente, en el 2013 se publicaron la mayor cantidad de artículos con un total de 76 (8.4%) teniendo así un mayor porcentaje entre todos los años, en el año 2014, 2015, 2016, 2017, 2018 se vio reflejado un promedio de 58 (6.5%) artículos publicados en los años mencionados; y en los años 2019, 2020, 2021 y 2022 se observaron un promedio de 64 (7%) artículos publicados en los 4 años.

**Tabla 6**

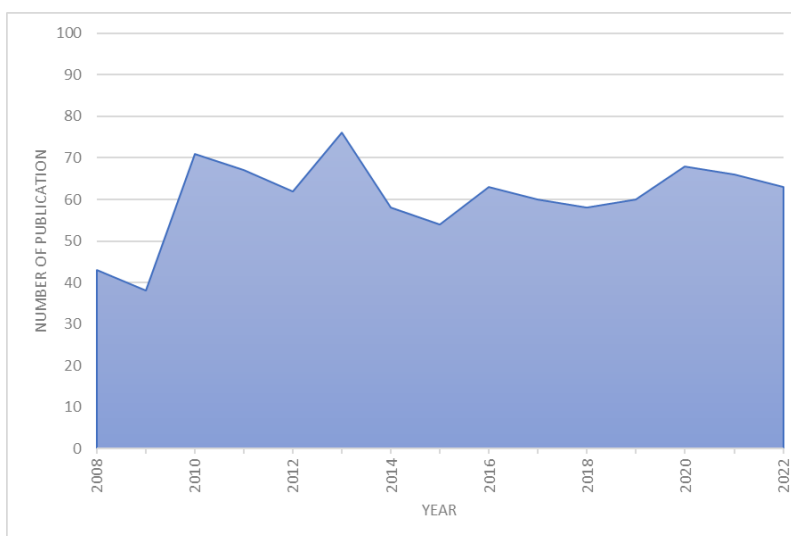
*Productividad anual de artículos*

<b>Año</b>	<b>Número de publicaciones</b>
2008	43
2009	38
2010	71
2011	67
2012	62
2013	76
2014	58
2015	54
2016	63
2017	60
2018	58
2019	60
2020	68
2021	66
2022	63

Para una mejor visualización de los datos (ver gráfico 1). En la figura se observa que existe un descenso en el año 2009, donde la curva muestra que en ese año solo se publicaron 38 artículos, mientras que en el año 2013 se observa un pico alto donde hubo un crecimiento significativo en donde se publicaron 76 artículos, teniendo como el mejor año en la cual se publicaron artículos mientras que en los años siguientes se observa curvas manteniéndose los artículos en promedio de 60 a 70 artículos publicados en los años siguientes.

**Gráfico 1**

*Crecimiento de producción anual*

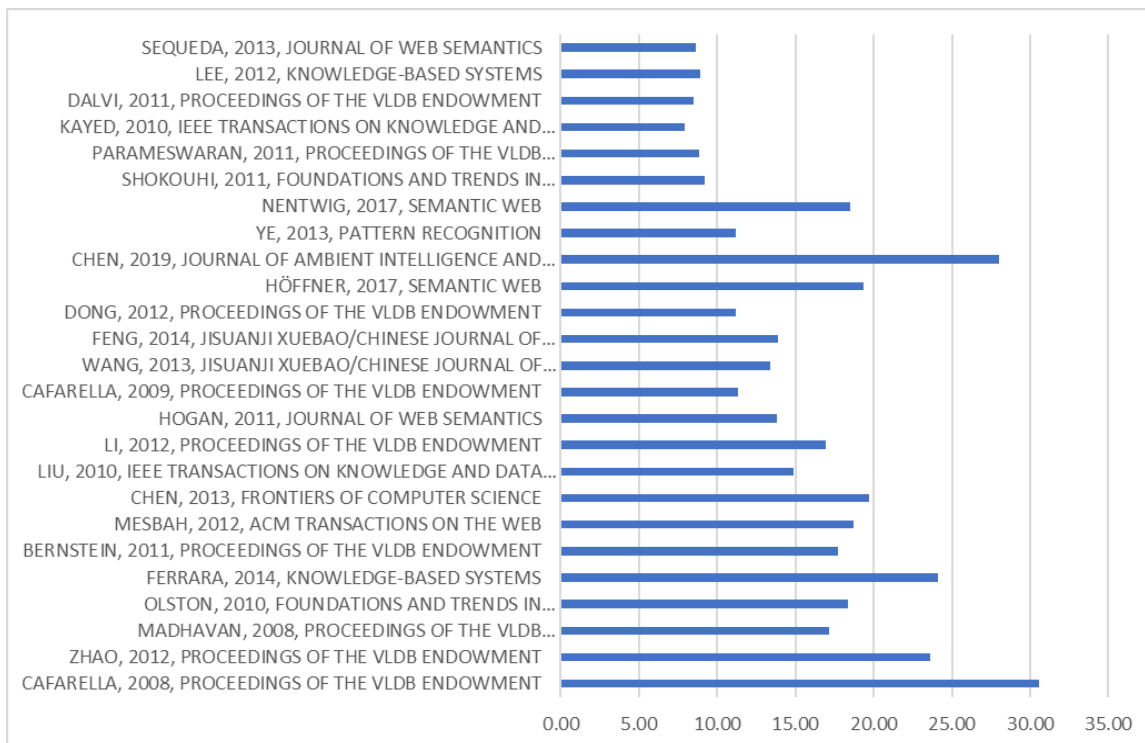


**Análisis de citas**

P7. ¿Cuáles son los artículos más referenciados acerca de la Deep Web en Ciencias Computacionales a nivel mundial? En la Tabla 7, se presentan los resultados de documentos más citados a nivel mundial sobre la Deep Web en Ciencias Computacionales, en lo cual se muestran CAFARELLA, 2008, PROCEEDINGS OF THE VLDB ENDOWMENT lidera con 459 citaciones en total (30.60 por año), ZHAO, 2012, PROCEEDINGS OF THE VLDB ENDOWMENT con 260 citaciones en total (23.60 por año), MADHAVAN, 2008, PROCEEDINGS OF THE VLDB ENDOWMENT con 257 citaciones en total (17.13 por año), siendo estos artículos los más referenciados en investigaciones, se puede decir que es un buen indicador para demostrar que los investigadores están produciendo gran cantidad de trabajos sobre la Deep Web o relacionados con ello, sin embargo, sería más óptimo que las citas sean sobre documentos más recientes. Para tener una visión general, en la gráfico 2, se grafica las tasas de crecimiento de producción científica por año de cada autor descrito en la Tabla 7.

**Gráfico 2**

*Artículos más citados a nivel mundial*

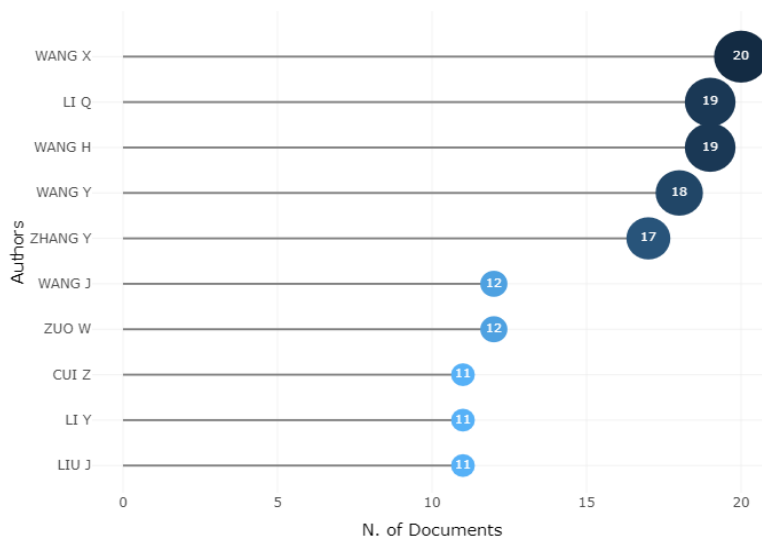


**Análisis de autores**

P8. ¿Qué autores son más relevantes en el campo de investigación sobre Deep web en las Ciencias Computacionales? El gráfico 3, representa los autores más relevantes en el campo de investigación sobre la Deep Web en Ciencias Computacionales en los últimos 15 años, en primer lugar, Wang X. (20 artículos), Li Q. y Wang H. (19 artículos), los 3 autores son de nacionalidad China.

**Gráfico 3**

La producción de los 10 autores más relevantes en el campo de investigación sobre Deep web en las Ciencias Computacionales



**Tabla 7**

Top 25 de los artículos más citados a nivel mundial

Artículo	DOI	Total de citaciones	TC por año	Normalized TC
CAFARELLA, 2008, PROCEEDINGS OF THE VLDB ENDOWMENT	10.14778/1453856.1453916	459	30.60	17.51
ZHAO, 2012, PROCEEDINGS OF THE VLDB ENDOWMENT	10.14778/2168651.2168656	260	23.64	12.55
MADHAVAN, 2008, PROCEEDINGS OF THE VLDB ENDOWMENT	10.14778/1454159.1454163	257	17.13	9.81
OLSTON, 2010, FOUNDATIONS AND TRENDS IN INFORMATION RETRIEVAL	10.1561/1500000017	239	18.38	14.87
FERRARA, 2014, KNOWLEDGE-BASED SYSTEMS	10.1016/j.knosys.2014.07.007	217	24.11	13.36
BERNSTEIN, 2011, PROCEEDINGS OF THE VLDB ENDOWMENT	-	213	17.75	12.14
MESBAH, 2012, ACM TRANSACTIONS ON THE WEB	10.1145/2109205.2109208	206	18.73	9.95
CHEN, 2013, FRONTIERS OF COMPUTER SCIENCE	10.1007/s11704-013-3903-7	197	19.70	11.03
LIU, 2010, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING	10.1109/TKDE.2009.109	193	14.85	12.01

LI, 2012, PROCEEDINGS OF THE VLDB ENDOWMENT	10.14778/2535568.2448943	186	16.91	8.98
HOGAN, 2011, JOURNAL OF WEB SEMANTICS	10.1016/j.websem.2011.06.004	166	13.83	9.46
CAFARELLA, 2009, PROCEEDINGS OF THE VLDB ENDOWMENT	10.14778/1687627.1687750	159	11.36	9.38
WANG, 2013, JISUANJI XUEBAO/CHINESE JOURNAL OF COMPUTERS	10.3724/SP.J.1016.2013.01125	134	13.40	7.50
FENG, 2014, JISUANJI XUEBAO/CHINESE JOURNAL OF COMPUTERS	10.3724/SP.J.1016.2014.00246	125	13.89	7.70
DONG, 2012, PROCEEDINGS OF THE VLDB ENDOWMENT	10.14778/2535568.2448938	123	11.18	5.94
HÖFFNER, 2017, SEMANTIC WEB	10.3233/SW-160247	116	19.33	11.07
CHEN, 2019, JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING	10.1007/s12652-018-01171-4	112	28.00	11.75
YE, 2013, PATTERN RECOGNITION	10.1016/j.patcog.2012.09.005	112	11.20	6.27
NENTWIG, 2017, SEMANTIC WEB	10.3233/SW-150210	111	18.50	10.59
SHOKOUHI, 2011, FOUNDATIONS AND TRENDS IN INFORMATION RETRIEVAL	10.1561/1500000010	110	9.17	6.27
PARAMESWARAN, 2011, PROCEEDINGS OF THE VLDB ENDOWMENT	10.14778/1952376.1952377	106	8.83	6.04
KAYED, 2010, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING	10.1109/TKDE.2009.82	103	7.92	6.41
DALVI, 2011, PROCEEDINGS OF THE VLDB ENDOWMENT	10.14778/1938545.1938547	102	8.50	5.81
LEE, 2012, KNOWLEDGE-BASED SYSTEMS	10.1016/j.knosys.2011.11.016	98	8.91	4.73
SEQUEDA, 2013, JOURNAL OF WEB SEMANTICS	10.1016/j.websem.2013.08.002	86	8.60	4.81

### **Análisis de palabras clave**

La tabla 8 muestra las palabras clave de autor más frecuentes y su frecuencia de aparición en los artículos relacionados con la Deep Web en Ciencias Computacionales. Las tres palabras clave más fuertes, "deep web", "ontology" y "semantic web" representan el alcance de la investigación, e indican el punto crítico de investigación y la frontera de estudio sobre la web profunda. Una nube de palabras puede ilustrar visualmente la situación de las palabras clave y resaltar las palabras clave con una alta frecuencia de ocurrencia. Para captar las palabras clave más destacadas en el campo de la web profunda de forma rápida e intuitiva, se generan nubes de palabras para las palabras clave extraídas del autor. La figura 6 muestra la nube de palabras claves de los autores. El tamaño de fuente de una palabra o frase representa su frecuencia de aparición. El gráfico 3 muestra la tendencia de variación de las 10 palabras clave de los autores más frecuentes en el campo de la web profunda en Ciencias

Computacionales. Las tendencias más representantes son "deep web" y "ontology" respectivamente, creciendo rápidamente en los últimos 15 años. Esta tendencia no solo indica que cada vez hay más publicaciones en el campo de estudio, sino que también señala el punto crítico de la investigación, de la internet oculta y el estudio de sus partes.

Figura 3

Nube de palabras de las palabras clave de autor más frecuente en publicaciones sobre la Deep Web en Ciencias Computacionales

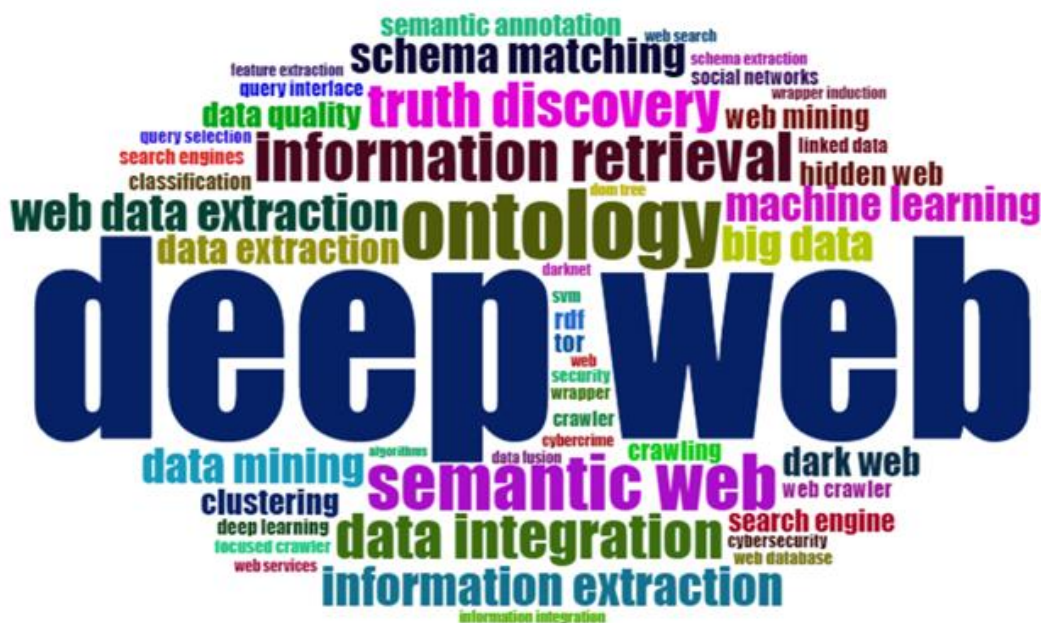


Tabla 8

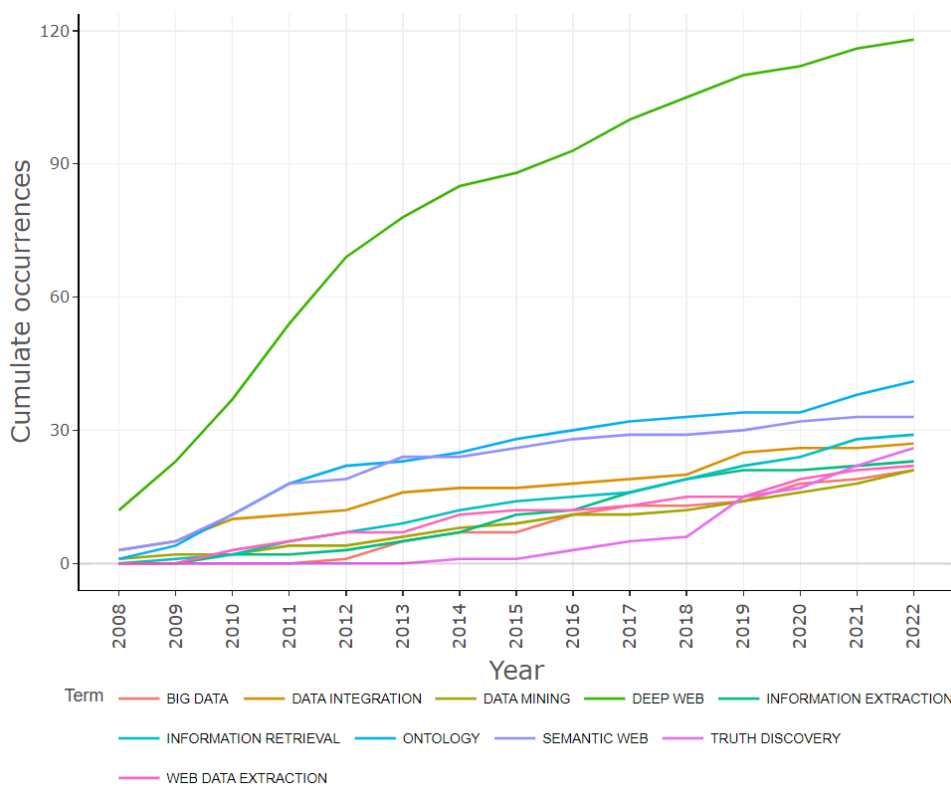
Las palabras claves de autores más frecuentes y su frecuencia de aparición en publicaciones sobre la Deep Web en Ciencias Computacionales

Términos	Frecuencia
deep web	118
ontology	41
semantic web	33
information retrieval	29
data integration	27
truth discovery	26
information extraction	23
web data extraction	22
big data	21
data mining	21
schema matching	21
machine learning	20
data extraction	18
dark web	17
clustering	15
data quality	15
hidden web	14

web mining	14
search engine	13
semantic annotation	13
crawling	12
rdf	12
tor	12
classification	10
web crawler	10

### Gráfico 3

La tendencia de variación (con Loess Smoothing) de las 10 palabras clave de autores más frecuentes en publicaciones sobre la Deep Web en Ciencias Computacionales



### Análisis Cualitativo

Las palabras clave más frecuentes se enumeran en la Tabla 8 y se dividen en categorías usando el método de escalamiento multidimensional, a partir de ello se generó el mapa de estructura conceptual, como se muestra en la Figura 4.

Las dimensiones del mapa representan la posición promedio de la publicación incluida en cada palabra clave, y el punto medio del mapa representa el centro del campo de investigación de la Deep Web en Ciencias Computacionales. Puede verse fácilmente en el mapa de estructura conceptual que las publicaciones sobre la internet profunda se dividen por colores y cada uno representa una categoría. El grupo naranja contiene la mayoría de las principales palabras clave, mientras que el grupo azul y marrón contienen las palabras clave un poco menos relevantes. Las publicaciones de estos grupos representan la línea de investigación sobre la Deep Web en el campo de las Ciencias Computacionales.





mantener así la investigación como un pilar de la educación. Por otra parte, los países tales como Argentina, Colombia, Perú, Venezuela, Ecuador, Brasil solo en el rango mencionado de los años 2008 y 2022 han ejercido sólo de 0 a 2 publicaciones, el país que más resalta es Chile con una cantidad de 4 artículos publicados, no están involucrados en el ámbito de la investigación. Por otro lado, se conocen las universidades e instituciones que lideran en el campo de investigación de la Deep Web.

Esta bibliometría está limitada primero por el número total de artículos de la base de datos seleccionada en este caso Scopus, contienen un gran volumen de artículos científicos y a pesar de trabajar solo con una pequeña parte, el análisis bibliométrico se dificulta por la cantidad de publicaciones seleccionadas (907 artículos). En segundo lugar, el rápido progreso en el campo de la tecnología y ciberseguridad y temas relacionados con la Deep Web y las Ciencias Computacionales, limita la oportunidad de realizar un análisis bibliométrico. En tercer lugar, enfatizamos que las actividades de investigación en este campo no necesariamente reflejan la aplicación práctica o el impacto de estos hallazgos de investigación. Los resultados de estos estudios sólo reflejan las tendencias de investigación actuales en la academia, que es el propósito de esta bibliometría.

### **CONCLUSIÓN**

En este estudio se realizó un análisis bibliométrico de los artículos publicados por los autores de las distintas universidades de todo el mundo entre 2008 y 2022. Se encontraron un total de 907 artículos a través de la base de datos Scopus. A nivel mundial, China, Estados Unidos, India y España son los países más destacados en este campo de investigación. También en este estudio, se encontró que Argentina, Colombia, Perú, Venezuela, Ecuador y Brasil han investigado poco o nada en esta área; por lo tanto, es tarea de estos países enfocarse en la investigación y motivar a otros investigadores universitarios a producir más documentos sobre la web profunda.

El análisis de palabras clave mostró que la web profunda es el foco de investigación en este campo en los últimos años. En la etapa de análisis cualitativo se analizaron 40 publicaciones con no menos de 10 índices de citas anuales.

Se aplicó el método de taxonomía para clasificar estas publicaciones se tomaron en cuenta las revisiones, algoritmos de segmentación y otras publicaciones relevantes. Este artículo es de gran importancia para comprender cómo la Deep Web ha adquirido mayor relevancia en las nuevas tendencias de las Ciencias Computacionales. Se espera que cada vez más investigadores trabajen en analizar la relación de la Deep Web en las Ciencias de la Computación ya que aún es un término no comprendido por muchos.

Como trabajo futuro, los nuevos artículos encontrados en la investigación, servirán para complementar el presente trabajo de investigación y posteriormente esperamos que estas puedan migrar a un repositorio netamente de temas referidos a la Deep Web.

## REFERENCIAS

- Bernstein, P. A., Madhavan, J., & Rahm, E. (2011). Generic schema matching, ten years later. *Proceedings of the VLDB Endowment*, 4(11), 695-701.
- Cafarella, M. J., Halevy, A., & Khoussainova, N. (2009). Data integration for the relational web. *Proceedings of the VLDB Endowment*, 2(1), 1090-1101.
- Cafarella, M. J., Halevy, A., Wang, D. Z., Wu, E., & Zhang, Y. (2008). Webtables: exploring the power of tables on the web. *Proceedings of the VLDB Endowment*, 1(1), 538-549.
- Camps D. Estudio bibliométrico general de colaboración y consumo de la información en artículos originales de la revista *Universitas Médica*, período 2002 a 2006, *Universitas Médica* 2007; 48(4): 358-365.
- Camps D. Limitaciones de los indicadores bibliométricos en la evaluación de la actividad científica biomédica, *Colombia Médica* 2008; 39(1): 74-79
- Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: a data management perspective. *Frontiers of computer Science*, 7(2), 157-164.
- Chen, Y., Wang, J., Xia, R., Zhang, Q., Cao, Z., & Yang, K. (2019). RETRACTED ARTICLE: The visual object tracking algorithm research based on adaptive combination kernel. *Journal of Ambient Intelligence and Humanized Computing*, 10(12), 4855-4867.
- Courset, F., Favennec, M., & Hamou, C. (2017). Un enfoque estratégico para la red Tor: *Revista de Inteligencia Estratégica de Amenazas*. Recuperado en septiembre de 2019, de <https://blogs.harvard.edu/cybersecurity/tag/deep-web/>
- Dalvi, N., Kumar, R., & Soliman, M. (2011). Automatic wrappers for large scale web extraction. *arXiv preprint arXiv:1103.2406*.
- Díaz, D. E. S. (2022). Análisis bibliométrico básico de la evolución científica de la deep web: web of science 2002-2021 (Doctoral dissertation, UNIVERSIDAD DE GRANADA).
- Dong, X. L., Saha, B., & Srivastava, D. (2012). Less is more: Selecting sources wisely for integration. *Proceedings of the VLDB Endowment*, 6(2), 37-48.
- Faruk, M., Rahman, M., & Hasan, S. (2021). How digital marketing evolved over time: A bibliometric analysis on scopus database. *Heliyon*, e08603.
- Feng, D., Zhang, M., & Li, H. (2014). Big data security and privacy protection. *Chinese Journal of computers*, 37(1), 246-258.
- Ferrara, E., De Meo, P., Fiumara, G., & Baumgartner, R. (2014). Web data extraction, applications and techniques: A survey. *Knowledge-based systems*, 70, 301-323.
- Franco, J. A. R. (2021). Desmitificando a la deep web a través de un fugaz viaje por la dark web. *Revista Ingeniería, Matemáticas y Ciencias de la Información*, 8(15), 13-32.
- Höffner, K., Walter, S., Marx, E., Usbeck, R., Lehmann, J., & Ngonga Ngomo, A. C. (2017). Survey on challenges of question answering in the semantic web. *Semantic Web*, 8(6), 895-920.
- Hogan, A., Harth, A., Umbrich, J., Kinsella, S., Polleres, A., & Decker, S. (2011). Searching and browsing linked data with swse: The semantic web search engine. *Journal of web semantics*, 9(4), 365-401.

Kayed, M., & Chang, C. H. (2010). FiVaTech: Page-level web data extraction from template pages. *IEEE transactions on knowledge and data engineering*, 22(2), 249-263.

Lee, M. R., & Chen, T. T. (2012). Revealing research themes and trends in knowledge management: From 1995 to 2010. *Knowledge-Based Systems*, 28, 47-58.

Li, Z., Xu, H., Guo, S., & Chen, K. (2012). Cloudvista: interactive and economical visual cluster analysis for big data in the cloud. *Proceedings of the VLDB Endowment*, 5(12), 1886-1889.

Liu, J., Li, J., Liu, C., & Chen, Y. (2010). Discover dependencies from data—a review. *IEEE Transactions on Knowledge and Data Engineering*, 24(2), 251-264.

Lobo, M. (2018). Un paseo por la Deep Web: Universitat Oberta de Catalunya. Recuperado en septiembre de 2019, de <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/72626/7/mloboromTFM0118memoria.pdf>

Madhavan, J., Ko, D., Kot, L., Ganapathy, V., Rasmussen, A., & Halevy, A. (2008). Google's deep web crawl. *Proceedings of the VLDB Endowment*, 1(2), 1241-1252.

Manzano Nunez, R., & García Perdomo, H. (2016). Sobre los criterios de inclusión y exclusión. Más allá de la publicación. *Revista Chilena de Pediatría*, 87(6), 511-512. doi:10.1016/j.rchipe.2016.05.003

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, 126(1), 871-906.

Medina López, C., Marín García, J., & Alfalla Luque, R. (2010). Una propuesta metodológica para la realización de búsquedas sistemáticas de bibliografía. 1(2), 13-30. doi:10.4995/wpom.v1i2.786

Mesbah, A., Van Deursen, A., & Lenselink, S. (2012). Crawling Ajax-based web applications through dynamic analysis of user interface state changes. *ACM Transactions on the Web (TWEB)*, 6(1), 1-30.

Molina-Muñoz, J. (2021). ANÁLISIS BIBLIOMETRICO DEL USO DE MACHINE LEARNING EN FINANZAS A TRAVÉS DE UN MODELO K-MEANS. *Revista Eficiencia*, 3(3).

Monroy, M. (2020) ¿Qué es la Deep Web y qué información podemos encontrar? *Publicación semestral*, Vol. 3, No. 5 (2020) 1-4

Nentwig, M., Hartung, M., Ngonga Ngomo, A. C., & Rahm, E. (2017). A survey of current link discovery frameworks. *Semantic Web*, 8(3), 419-436.

Olston, C., & Najork, M. (2010). Web crawling. *Foundations and Trends® in Information Retrieval*, 4(3), 175-246.

Parameswaran, A., Sarma, A. D., Garcia-Molina, H., Polyzotis, N., & Widom, J. (2011). Human-assisted graph search: it's okay to ask questions. *arXiv preprint arXiv:1103.3102*.

Pérez, D. (2018). Las 10 cosas más impactantes de la Deep Web. Recuperado en septiembre de 2019, de <https://www.linkedin.com/pulse/las-10-cosas-mas-impactantes-dela-deep-web-catal%C3%A1n-p%C3%A9rez-salazar>

Ramírez, P. E., Mariano, A. M., & Salazar, E. A. (2014). Propuesta Metodológica para aplicar modelos de ecuaciones estructurales con PLS: El caso del uso de las bases de datos científicas en estudiantes universitarios. *Revista ADMpg*, 7(2).

Sánchez-Céspedes, J. M., Rodríguez-Miranda, J. P., & Salcedo-Parra, O. J. (2020). Análisis de la producción de publicaciones científicas en inteligencia artificial aplicada a la formulación de Políticas Públicas. *Revista Científica*, 39(3), 353–368. <https://doi.org/10.14483/23448350.16301>

Segura, Elena. (2021). Análisis bibliométrico básico de la evolución científica de la deep web: web of science 2002-2021 Universidad de Granada

Semaan-Llurba S. (2018). ¿Qué es Scopus? ¿Y para qué sirve? Biblioteca Sant Joan de Déu. Disponible en <https://bibliosjd.org/2018/01/24/scopus-que-es-para-que-sirve/#.Y2XThnbMK3A>

Sequeda, J. F., & Miranker, D. P. (2013). Ultrawrap: SPARQL execution on relational data. *Journal of Web Semantics*, 22, 19-39.

Shokouhi, M., & Si, L. (2011). Federated search. *Foundations and trends® in information retrieval*, 5(1), 1-102.

Toro, A. (2018). La web profunda, un sitio entre sombras y realidades. Universidad de Manizales. Facultad de Ciencias e Ingeniería. <https://doi.org/10.30554/ventanainform.39.3311.2018>

Villada, D. & Jiménez, A. (2017). La Web Semántica y la Web Profunda como Sistemas de Información: Análisis a una realidad. *Revista Antioqueña de las Ciencias Computacionales y la Ingeniería de Software, RACCIS*, Vol. 7, No. 1. Medellín (Colombia): Instituto Antioqueño de Investigación, IAI. p. 43-51. ISSN: 2248-7441. <http://fundacioniai.org/raccis/v7n1/n12a4.pdf>

Wang, G., & Gui, X. L. (2013). Selecting and trust computing for transaction nodes in online social networks. *Jisuanji Xuebao(Chinese Journal of Computers)*, 36(2), 368-383.

Ye, Y., Wu, Q., Huang, J. Z., Ng, M. K., & Li, X. (2013). Stratified sampling for feature subspace selection in random forests for high dimensional data. *Pattern Recognition*, 46(3), 769-787.

Zhao, B., Rubinstein, B. I., Gemmell, J., & Han, J. (2012). A bayesian approach to discovering truth from conflicting sources for data integration. *arXiv preprint arXiv:1203.0058*.