

# Base De Datos

## I. RESUMEN

Para esta base de datos, se realizó una consulta inicial de 18,645 artículos relacionados con el tema de ('Frameworks' AND 'Web Development'). Decidi elegir este tema ya que los frameworks son componentes esenciales en el desarrollo de software, proporcionando una base estructurada que permite crear aplicaciones de manera eficiente siendo usadas en nuestros proyectos de software. Al ofrecer una amplia variedad de herramientas predefinidas y establecer convenciones de codificación, los frameworks simplifican tareas comunes y ayudan a mantener la coherencia en el código base. Posteriormente, se lleva a cabo un análisis de estos artículos para identificar los más relevantes en relación con el tema elegido. Este proceso de análisis permite seleccionar aquellos artículos que mejor se ajustan a los criterios establecidos, proporcionando así una buena base para la investigación o la búsqueda de información específica en el ámbito del desarrollo web y frameworks. Query Inicial: TITLE-ABS-KEY (( 'frameworks' AND 'web AND development' ) . )

## II. INTRODUCCIÓN

Esta investigación se llevó a cabo con el objetivo de explorar y analizar diversos artículos sobre el tema de frameworks y el desarrollo web, utilizando una metodología que combina el análisis de datos y la revisión bibliográfica. Para la recopilación de artículos, se utilizó la plataforma Scopus, reconocida por su extensa base de datos de investigación científica. Para ello, se realizó una búsqueda utilizando términos clave relacionados con framework y desarrollo web. Limitando la búsqueda de artículos con temas específicos y disminuyendo la cantidad de artículos a analizar resultó en una selección inicial de 560 artículos pertinentes.

Una vez recopilados, los artículos se exportaron en formato CSV y se trasladaron al entorno de Google Colab, para el análisis de datos. Utilizando el código preestablecido proporcionado en Python por el instructor, se llevó a cabo un análisis detallado de los artículos descargados, con el objetivo de identificar aquellos más relevantes y significativos para el tema de interés.

## III. RESULTADOS

1. Se empieza con un conjunto de importaciones en Python que trae herramientas para análisis y procesamiento de texto. La biblioteca **Pandas** facilita la manipulación de datos tabulares, mientras que **NLTK** ofrece una funciones para procesamiento de lenguaje natural, como la tokenización de oraciones y palabras, y la eliminación de stopwords. La función **matplotlib.pyplot** permite la creación de visualizaciones como gráficos, y **WordCloud** es especialmente útil para generar nubes de palabras que resalten las más frecuentes en un texto.
2. Posteriormente descargando el conjunto de palabras vacías (stopwords) de la biblioteca **NLTK**. Estás palabras comunes se eliminan del análisis de texto ya que en su mayoría no aportan significado contextual, como "a", "el", "en", etc. Así se filtran mejor y mejora la calidad del análisis de texto.
3. Con la siguiente función de **nltk.download('punkt')** se descarga el modelo de datos necesario para utilizar el tokenizador de oraciones y palabras proporcionado por **NLTK**.
4. Se utiliza la biblioteca **Pandas** para leer el archivo CSV llamado '**scopus.csv**' y cargar su contenido en el DataFrame. Luego, el método **head()** se utiliza para mostrar las primeras filas del DataFrame, lo que proporciona una vista previa de los datos y cómo estos están estructurados.

	Authors	Author full name	Author(s) ID	Title	Year	Source title	Link	Abstract	Author Keywords	Index Keywords	Open Access
1	Pathak V, Shrivastava V., Patel R, Gupta S,	Pathak, Vishal (57290577800); Shrivastava, Vikram (57200996000); Patel, Rajesh (57219434638); Gupta, Smita (57140000000)	57204854160;	Current Web Development Trends in Web and Systems	2023	Lecture Notes in Information Systems and	<a href="https://www.scopus.com/inward/record.uri?eid=2-s2.0-57204854160&amp;partnerID=40">https://www.scopus.com/inward/record.uri?eid=2-s2.0-57204854160&amp;partnerID=40</a>	This paper highlights the current trends in web development. It discusses various aspects of web development such as front-end development, back-end development, and mobile web development. It also covers emerging technologies like AI, machine learning, and blockchain in web development.	front-end development	back-end development	No
2	Chen Y., Lu Y., Peng F., Chen J., Shi F., Li Y.,	Chen, Yuanchao (57219434638); Lu, Liang (57200996000); Peng, Xiaoyang (57200996000); Chen, Jian (57200996000); Shi, Fuxiang (57200996000); Li, Yan (57200996000)	57219434638;	APIs for Identifying Applications on Web	2024	Electronics (Switzerland)	<a href="https://www.scopus.com/inward/record.uri?eid=2-s2.0-57219434638&amp;partnerID=40">https://www.scopus.com/inward/record.uri?eid=2-s2.0-57219434638&amp;partnerID=40</a>	Modern web technologies offer various APIs for...	modern web technologies	similarities between various web APIs...	No
3	Gedira G., Mitra M., Elouadi H., K.M. Ponnam J.G	Gedira, Ghassen (57219434638); Mitra, Monisha (57200996000); Elouadi, Hassan (57200996000); K.M. Ponnam, J.G (57200996000)	57219434638;	Integrations and Applications of Angular Framework	2022	3rd International Conference on Power, Energy,	<a href="https://www.scopus.com/inward/record.uri?eid=2-s2.0-57219434638&amp;partnerID=40">https://www.scopus.com/inward/record.uri?eid=2-s2.0-57219434638&amp;partnerID=40</a>	In today's world of ever-evolving technologies...	Angular; AngularJS; Client-side Routing; Front-end Frame...	HTML; Angular; AngularJS; Client-side Routing; Front-end Frame...	No
4	Vegapantam, J., Hettiarachchi, A.	Vegapantam, Julio (57022015930); Hettiarachchi, Anuradha (57192188480)	57022015930;	The State of Disappearing Frameworks	2023	International Conference on Web Disappearing frameworks	<a href="https://www.scopus.com/inward/record.uri?eid=2-s2.0-57022015930&amp;partnerID=40">https://www.scopus.com/inward/record.uri?eid=2-s2.0-57022015930&amp;partnerID=40</a>	Disappearing frameworks	Auto-Code Splitting;	High-level languages; Astro; React;	All Open Access; Gold Open Access

Figura 1. Estructuración de los datos.

5. Luego de ello, se crea un conjunto de palabras vacías (stopwords) en inglés utilizando la biblioteca **NLTK**. La función `stopwords.words('english')` carga un conjunto predefinido de stopwords en inglés, y luego este conjunto se convierte en un conjunto de Python utilizando la función `set()`. Una vez creado, este conjunto puede ser utilizado para filtrar stopwords de un texto en inglés durante el procesamiento de lenguaje natural.

6. Después se seleccionan 150 títulos más frecuentes agrupados por la columna 'Source title' del DataFrame, que contiene los datos de los artículos obtenidos de la base de datos Scopus.

Ya teniendo esto se crea una visualización utilizando Matplotlib. Se especifica un tamaño de figura de 20 por 5 pulgadas con la función `(plt.figure(figsize=(20, 5)))`. Por otra parte, Se utiliza la función `kind='bar'` en el método `plot()` para generar un gráfico de barras que muestra la cantidad de títulos para cada fuente. Finalmente, `plt.show()` se utiliza para mostrar la visualización.

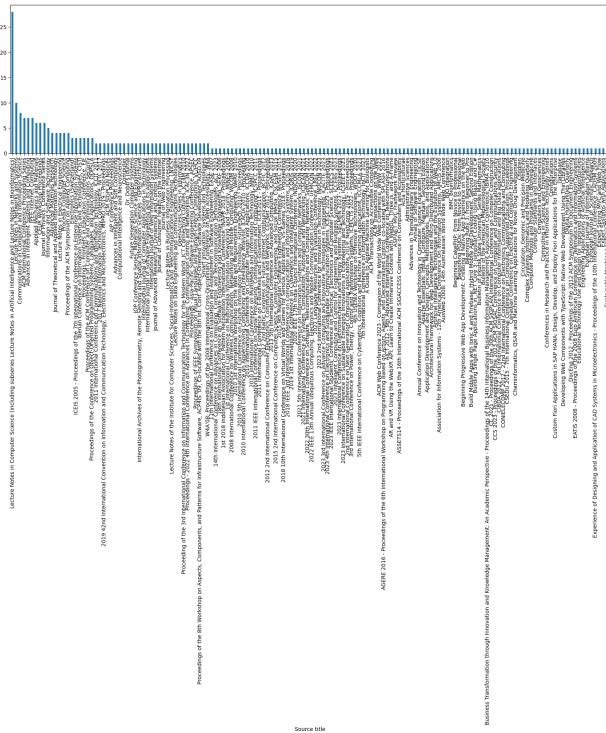


Figura 2. Agrupación de artículos por título.

7. Se agrupan los datos por el título de la fuente ('Source title') y cuenta el número de artículos de cada fuente. Luego, selecciona las 10 fuentes con más artículos y crea

un gráfico de barras que muestra el número de artículos publicados por cada una de estas fuentes.

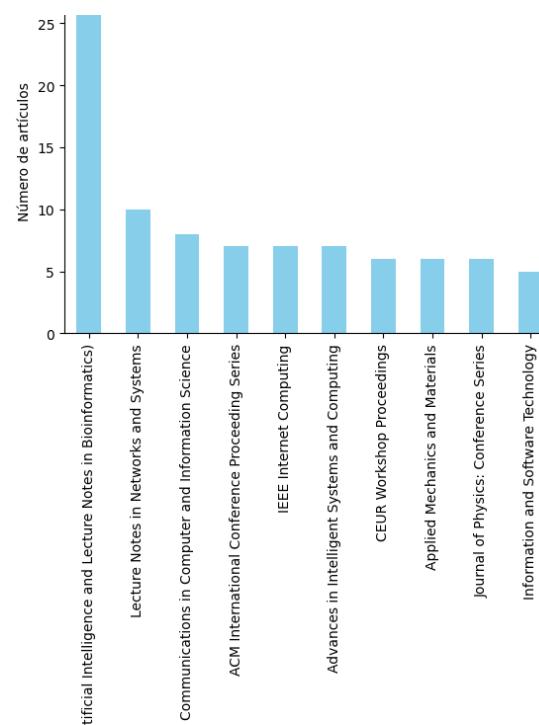


Figura 3. Fuentes con más artículos.

8. Se usa La función `clean_text` la cual prepara y limpia el texto antes de realizar un análisis de texto o procesamiento de lenguaje natural (NLP). Para ello se elimina la puntuación del texto, como caracteres irrelevantes que no aportan significado al análisis. Además, filtra las palabras vacías, conocidas como stopwords.
9. Posteriormente se realiza un análisis de los títulos de los artículos en el DataFrame, con el objetivo de identificar las palabras más frecuentes y relevantes en el de texto. Para ello se cuenta la frecuencia de cada palabra en todos los títulos y crea un DataFrame que contiene estas palabras junto con su frecuencia de aparición. Este DataFrame se ordena según la frecuencia de las palabras de mayor a menor, y se le asigna a cada palabra un rango numérico en función

de su frecuencia. De esta manera se ofrece una visión rápida de las palabras más comunes en los títulos de los artículos.

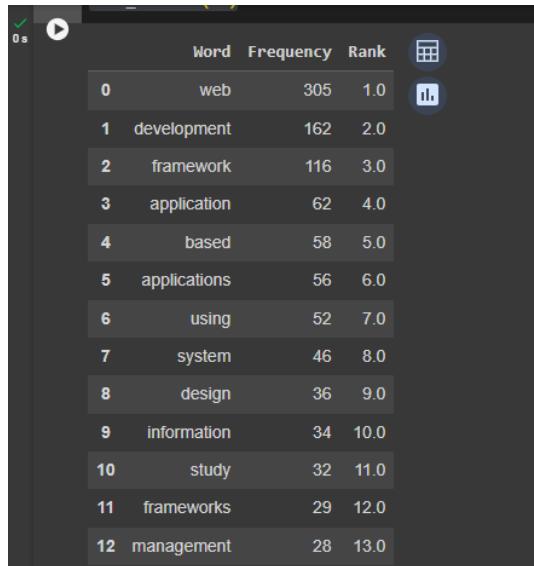


Figura 4. Ranking De Palabras.

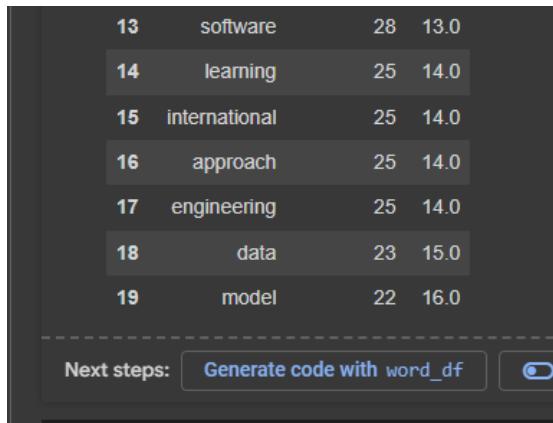


Figura 5. Ranking De Palabras.



Figura 6. Nube De Palabras.

11. Se utiliza la biblioteca Matplotlib para crear un gráfico de barras que muestra las 25 palabras más frecuentes en el DataFrame. Para ello, se ordena el DataFrame por la frecuencia de las palabras en orden descendente y se seleccionan las 25 palabras más frecuentes. Luego, se utiliza el método `plot()` del DataFrame para generar el gráfico de barras, especificando las columnas 'Word' y 'Frequency', y se indica que se desea un gráfico de barras con la función `kind='bar'`. Esto proporciona una representación concisa de las palabras más frecuentes en el DataFrame, lo que permite identificar las principales palabras clave.

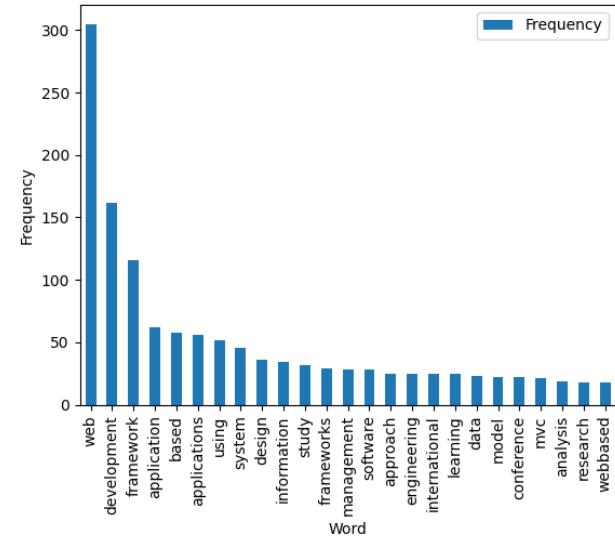


Figura 7. Palabras más frecuentes.

10. Ya teniendo el ranking de palabras se genera una gráfica con una nube de palabras a partir de los títulos de los artículos, lo que proporciona una vista rápida de las palabras más frecuentes en el texto. Para lograr esto, se aplica una función de limpieza de texto a cada título. Luego, concatena todos los títulos limpios en un solo texto. Finalmente, utilizando la biblioteca **WordCloud**, se crea una nube de palabras donde el tamaño de cada palabra está determinado por su frecuencia en el texto. Es decir, las palabras más frecuentes son más prominentes, lo que permite identificar rápidamente los temas dominantes en los títulos de los artículos.
12. Teniendo en cuenta lo anterior se realiza un análisis detallado de los títulos de los artículos, enfocándose en identificar aquellos que contienen un mayor número de palabras clave. Es importante hacer uso de la función `clean_text` ya que limpia los títulos y utilizando un objeto Counter, se cuentan las frecuencias de las palabras clave en los títulos, así se seleccionan las 15 palabras clave más comunes para el análisis. Luego, se define una función `count_keywords` para determinar cuántas de estas palabras clave están presentes en cada título de artículo. Esto permite identificar los más relevantes en función de la

presencia de estas palabras clave. Finalmente, se visualizan los 10 principales artículos con la mayor cantidad de palabras clave a través de un gráfico de barras.

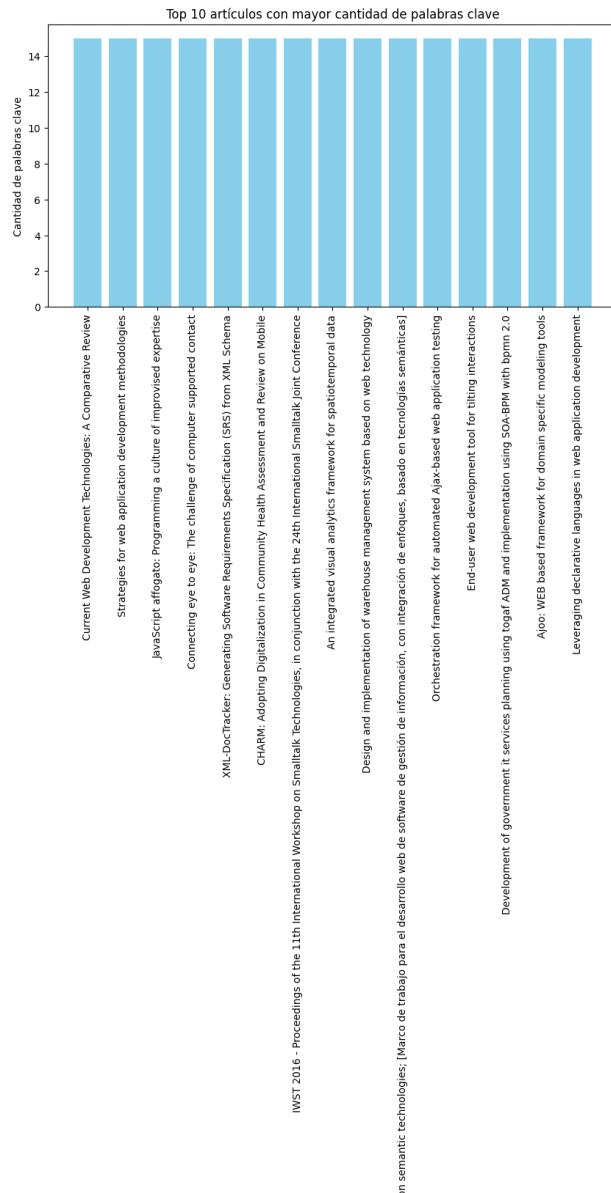


Figura 8. Artículos con más palabras clave.

13. Ya teniendo los artículos más relevantes se resumen los abstracts de los artículos que contienen un mayor número de palabras clave. Para ello, utiliza el DataFrame previamente ordenado por el recuento de palabras clave para seleccionar los abstracts de los 10 artículos principales. Luego, define una función llamada *summarize\_abstract* que toma un abstract como entrada y lo resume en tres oraciones clave. Esto se logra mediante el cálculo de una puntuación para cada oración. Posteriormente, aplica esta función de resumen a cada abstract seleccionado utilizando el método *apply*, generando así una serie de

resúmenes para los abstracts de los artículos seleccionados. Finalmente, muestra cada resumen del 1 al 10 para indicar su posición en la lista de abstracts seleccionados.

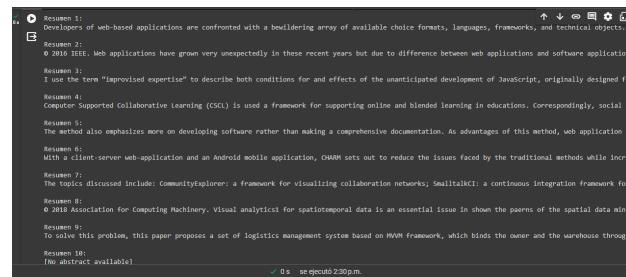


Figura 9. Resumen de los 10 artículos.

14. Al identificar los artículos más relevantes mediante el conteo de palabras clave tanto en sus títulos como en sus abstracts. Se cuenta la aparición de palabras clave en cada título de artículo, utilizando una función *lambda* que aplica la limpieza de texto y luego suma el número de palabras clave presentes en el título. Luego, se seleccionan los abstracts de los 10 artículos con mayor conteo de palabras clave. Se hace un resumen con la función *summarize\_abstract* se aplica a cada abstract seleccionado y se almacena junto con el título correspondiente en el DataFrame *top\_articles*. Finalmente, se muestra cada título de artículo junto con su resumen correspondiente.

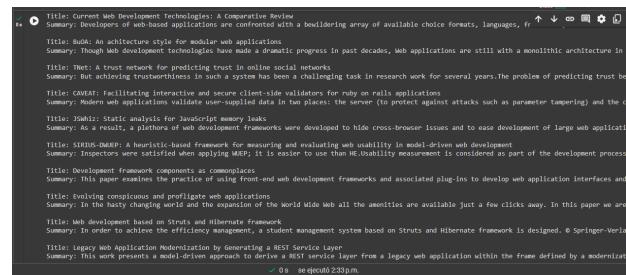


Figura 10. Artículos más relevantes.

#### IV. QUERY FINAL

Al momento de exportar el CSV se limitó la búsqueda de artículos por lo que a continuación se muestra el query dado en excel al momento de exportar el artículo. Query Final: Authors,"Author full names","Author(s) ID","Title","Year","Source title","Link","Abstract","Author Key-words","Index Keywords","Open Access".

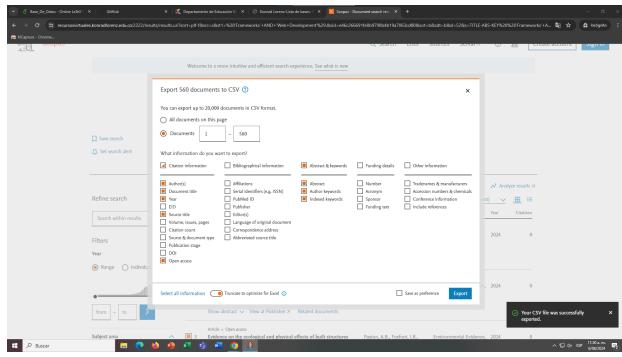


Figura 11. Limitación de artículos.

## V. ANÁLISIS COMPARATIVO

Se realiza un análisis comparativo de los resultados obtenidos en la investigación sobre 'frameworks' y 'web development' en diferentes períodos, como los últimos 5 años frente a los 5 años anteriores a esos, permite identificar cambios significativos en las tendencias y enfoques de investigación. Este análisis revela cómo han evolucionado las prioridades y las metodologías en el campo. Además de ello no solo destaca los cambios en los enfoques metodológicos y las áreas de interés predominantes, si no que se ha proporcionando una visión clara de cómo el campo ha respondido a nuevas demandas. Para relaizar este analisis comparativo primero se analizará el periodo de tiempo entre el 2018 al 2024 y posteriormente entre 2013 al 2018.

## VI. RESULTADOS- PRIMEROS 5 AÑOS 2018 - 2024

- Los primeros 5 años se limitaron al rango desde el 2018 al 2024, este periodo comprendido entre 2018 y 2024 ha sido testigo de avances significativos en el ámbito del desarrollo web. Durante estos años, los frameworks han desempeñado un papel crucial en la simplificación y aceleración del desarrollo de aplicaciones web, proporcionando herramientas para manejar la complejidad creciente de las aplicaciones modernas. Query Inicial: TITLE-ABS-KEY ( "frameworks" AND "web development" ) AND PUBYEAR > 2017 AND PUBYEAR < 2025.

1. Entre 2018 y 2024, el desarrollo de frameworks web ha experimentado avances significativos, reflejando cambios en las necesidades y enfoques de la industria del desarrollo web. Como en el anterior ejercicio, este análisis utiliza herramientas de procesamiento de texto y análisis de datos como Pandas, NLTK, Matplotlib y WordCloud para examinar publicaciones académicas y técnicas en Scopus. Estas herramientas permiten manipular datos tabulares, procesar lenguaje natural y visualizar información clave. A continuación, se presenta la estructura de los datos

basada en cuatro artículos seleccionados de ejemplo. En este caso, se puede observar los artículos de los últimos años de manera estructurada.

Category	Authors	Author's full names	Author(s) ID	Title	Year	Source title	Link	Abstract	Author Key words	Index Keys	Open Access
0	Fatih V., Srinivasan B., But R., Gupta A.V.S.S., Chukwuemeka O., Kothiyal R., Chen Y., Pan Z., Geetha G., Muralidharan P., Prasad K.M., Persson J.G., Venkatesan H., J. Helas A., Vuorimaa P.	Patilak, Vishwanath (57264054165); Pathak, Vibhav (57264054166); Pathak, Vibhav (5688077000); Pathak, Vibhav (5684736700); Somvanshi, A.V.S.S. (57264054167); Somvanshi, A.V.S.S. (5689167900); Chen, Yuanchao (57264054168); Chen, Yuanchao (57264054169); Li, Xiangyu (57264054170); Li, Xiangyu (567263); Geetha G. (57212263280); Muralidharan P. (57212263281); Prasad K.M. (576); Persson J.G. (57292158000); Venkatesan H. (57292158001); J. Helas A. (5682015900); Vuorimaa P. (660324155)	Patilak, Vibhav (57264054165); Pathak, Vibhav (57264054166); Pathak, Vibhav (5688077000); Pathak, Vibhav (5684736700); Somvanshi, A.V.S.S. (57264054167); Somvanshi, A.V.S.S. (5689167900); Chen, Yuanchao (57264054168); Chen, Yuanchao (57264054169); Li, Xiangyu (57264054170); Li, Xiangyu (567263); Geetha G. (57212263280); Muralidharan P. (57212263281); Prasad K.M. (576); Persson J.G. (57292158000); Venkatesan H. (57292158001); J. Helas A. (5682015900); Vuorimaa P. (660324155)	Current Web Technologies; A 2023 (2023-01-01)	2023	Lecture Notes in Intelligent Systems and Applications	<a href="https://www.scopus.com/inward/record.uri?eid=2-s2.0-53000000000&amp;partnerID=M4RI">https://www.scopus.com/inward/record.uri?eid=2-s2.0-53000000000&amp;partnerID=M4RI</a>	This paper highlights current web technologies.	Web-based; Back-end and front-end development	Web-based; Back-end and front-end development	No
1				MemoryDB Integration with Python and Java	2024	4th International Conference on Advances in Computing and Mathematics	<a href="https://www.scopus.com/inward/record.uri?eid=2-s2.0-53000000000&amp;partnerID=M4RI">https://www.scopus.com/inward/record.uri?eid=2-s2.0-53000000000&amp;partnerID=M4RI</a>	This scholarly article presents a detailed analysis.	MemoryDB; NSQL; Integration; Application Programming (API); High-level language	Application Programming (API); High-level language	No
2				Identifying Web Application APIs from URLs	2024	Electronics (Switzerland)	<a href="https://www.scopus.com/inward/record.uri?eid=2-s2.0-53000000000&amp;partnerID=M4RI">https://www.scopus.com/inward/record.uri?eid=2-s2.0-53000000000&amp;partnerID=M4RI</a>	Modern web applications offer various features.	similarity; static; state-aware; web; API; web	All Open Access Content; Open Access	No
3				Interpretation and Analysis of a New Framework	2023	3rd International Conference on Power Energy	<a href="https://www.scopus.com/inward/record.uri?eid=2-s2.0-53000000000&amp;partnerID=M4RI">https://www.scopus.com/inward/record.uri?eid=2-s2.0-53000000000&amp;partnerID=M4RI</a>	In today's world of rapid energy technology.	Angular; React; Client-side Rendering; Framework	HTML; Angular; React; Client-side Rendering; Framework	No
4				Disappearing Frameworks: Is there a new type o...	2023	International Conference on Web Information Systems	<a href="https://www.scopus.com/inward/record.uri?eid=2-s2.0-53000000000&amp;partnerID=M4RI">https://www.scopus.com/inward/record.uri?eid=2-s2.0-53000000000&amp;partnerID=M4RI</a>	Disappearing frameworks are becoming a new type o...	Astro; Code Splicing; Disappearing Frameworks	High level; language; Open source; Astro; Code Splicing; Disappearing Frameworks	All Open Access Content; Open Access

Figura 12. Estructuración de los datos.

2. Para el análisis de datos obtenidos de Scopus, se emplea la biblioteca NLTK para crear un conjunto de palabras vacías (stopwords) en inglés, utilizando la función stop\_words.words('english'). A continuación, se seleccionan los 150 títulos más frecuentes agrupados por la columna 'Source title' del DataFrame.

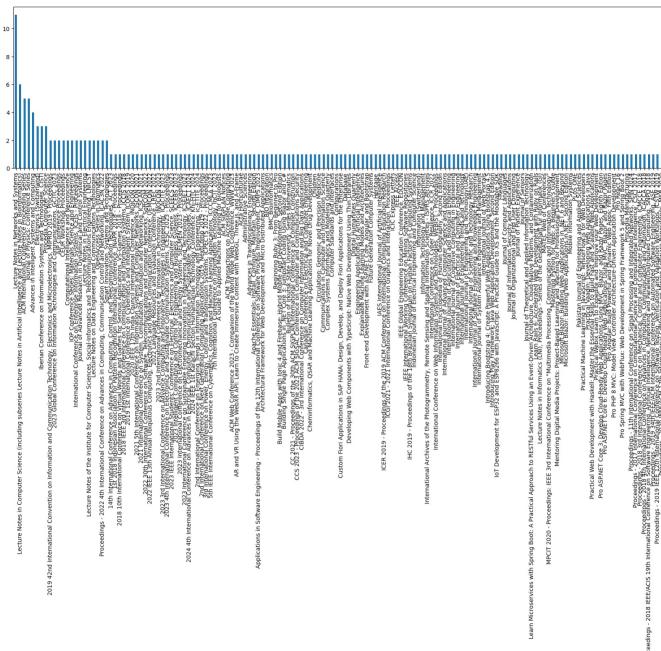


Figura 13. Agrupación de artículos por título.

3. Se agrupan los datos por el título de la fuente ("Source Title") y se cuenta el número de artículos de cada fuente. Luego, se seleccionan las 10 fuentes con más artículos y se crea un gráfico de barras que muestra el número de artículos publicados por cada una de estas fuentes. En este caso, la fuente con más artículos es "*Lecture Notes in Networks and Systems*" y la que tiene menos es "*OpenAccess Series in Informatics*".

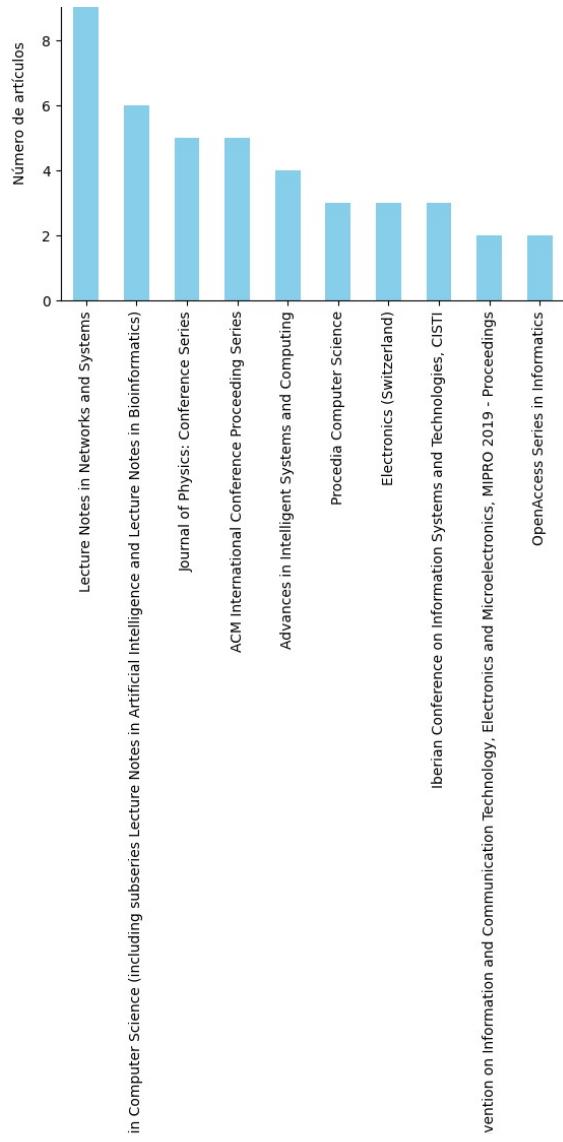


Figura 14. Fuentes con más artículos.

Word	Frequency	Rank
0 web	127	1.0
1 development	75	2.0
2 framework	49	3.0
3 using	39	4.0
4 application	37	5.0
5 based	34	6.0
6 system	26	7.0
7 applications	26	7.0
8 learning	21	8.0
9 design	19	9.0
10 frameworks	19	9.0
11 study	19	9.0

Figura 15. Ranking De Palabras.

12	data	17	10.0
13	approach	16	11.0
14	de	13	12.0
15	analysis	12	13.0
16	python	11	14.0
17	javascript	11	14.0
18	software	11	14.0
19	mvc	11	14.0

Figura 16. Ranking De Palabras.

4. Para llevar a cabo un análisis detallado de los títulos de los artículos, se utiliza la función `clean_text`, que limpia el texto antes de realizar cualquier análisis o procesamiento de lenguaje natural (NLP). Posteriormente, se analizan los títulos de los artículos en el DataFrame para identificar las palabras más frecuentes y significativas. La frecuencia de cada palabra se cuenta en todos los títulos y se crea un DataFrame que contiene estas palabras junto con su frecuencia de aparición. Este DataFrame se ordena de mayor a menor frecuencia y se asigna a cada palabra un rango numérico en función de su frecuencia. En este caso la palabra con más frecuencia es "web" y la menor es "mvc".

5. Ya teniendo el ranking de palabras se genera una gráfica con una nube de palabras a partir de los títulos de los artículos, lo que proporciona una vista rápida de las palabras más frecuentes en el texto. Para lograr esto, se aplica una función de limpieza de texto a cada título. Luego, concatena todos los títulos limpios en un solo texto. Finalmente, utilizando la biblioteca **WordCloud**, se crea una nube de palabras donde el tamaño de cada palabra está determinado por su frecuencia en el texto. En este caso, se pueden observar algunas palabras más sobresalientes como web development ,application ,framework, using web entre otras.

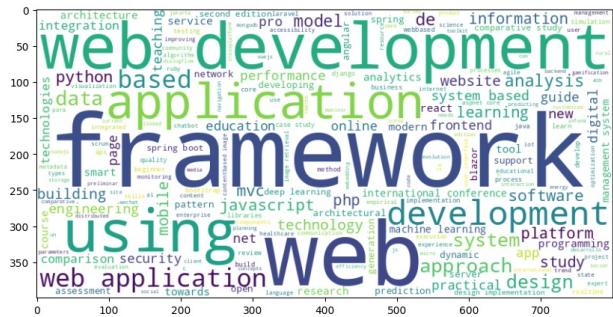


Figura 17. Nube De Palabras.

6. Se utiliza la biblioteca Matplotlib para crear un gráfico de barras que muestra las 25 palabras más frecuentes en el DataFrame. Para ello, se ordena el DataFrame por la frecuencia de las palabras en orden descendente y se seleccionan las 25 palabras más frecuentes. Luego, se utiliza el método `plot()` del DataFrame para generar el gráfico de barras, especificando las columnas 'Word' y 'Frequency', y se indica que se desea un gráfico de barras con la función `kind='bar'`. Como se puede observar la palabra más frecuente es "web" y la menos frecuente es "information".

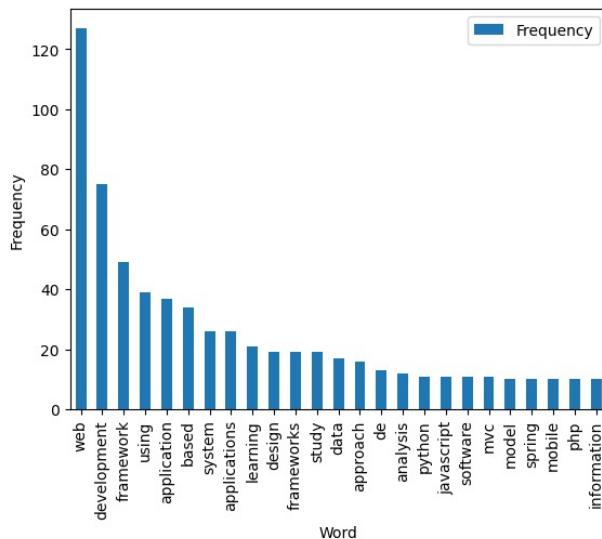


Figura 18. Palabras más frecuentes.

7. Teniendo en cuenta lo anterior se realiza un análisis detallado de los títulos de los artículos, enfocándose en identificar aquellos que contienen un mayor número de palabras clave. Es importante hacer uso de la función `clean_text` ya que limpia los títulos. Luego, se define una función `count_keywords` para determinar cuántas de estas palabras clave están presentes en cada título de los 10 artículos. En esta imagen podemos observar que el artículo con más palabras clave es "Current Web Development Technologies A Comparative Review" y el menor es "Research and application of template engine for web back-end based on Mybatis-PLUS".

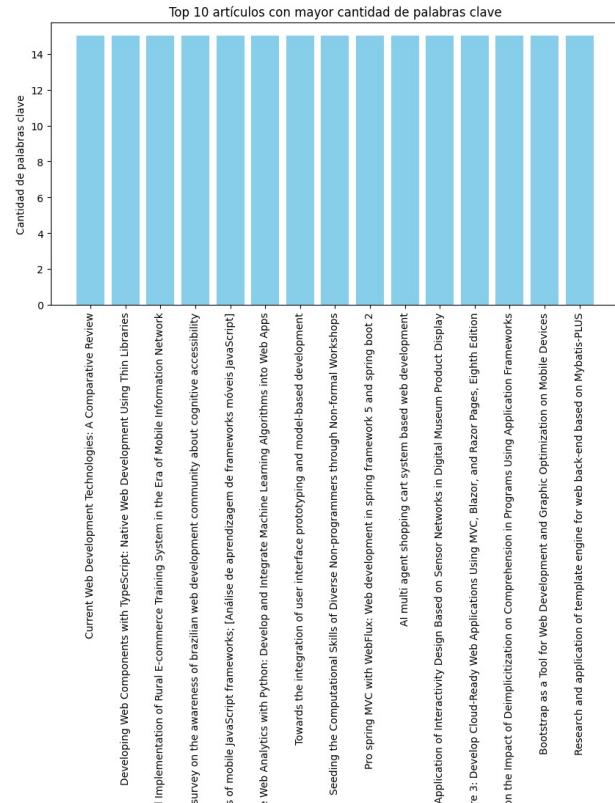


Figura 19. Artículos con más palabras clave.

8. Ya teniendo los artículos más relevantes se resumen los abstracts de los artículos que contienen un mayor número de palabras clave. Para ello, utiliza el DataFrame previamente ordenado por el recuento de palabras clave para seleccionar los abstracts de los 10 artículos principales. Luego, define una función llamada `summarize_abstract` que toma un abstract como entrada y lo resume en tres oraciones clave. Esto se logra mediante el cálculo de una puntuación para cada oración. Finalmente, muestra cada resumen del 1 al 10 para indicar su posición en la lista de abstracts seleccionados.

```

Resumen 1:
Developers of web-based applications are confronted with a bewildering array of available choice formats, languages, frameworks, and technical objects. The confusion of technology is
Resumen 2:
What will learn + Create isolated web components using shadow DOM, slots, and templates + Understand the advantage of an enhanced toolset, especially TypeScript + Pick up styles
Resumen 3:
Instead, I am exploring three blazes that potentially prevent those professionals from approaching cognitive disabilities: strong organizational barriers; difficulty to understand its
Resumen 4:
The results evidenced three blazes that potentially prevent those professionals from approaching cognitive disabilities: strong organizational barriers; difficulty to understand its
Resumen 5:
This paper is currently a popular language mainly in the creation of client-side code for both web development and, more and more, for the mobile application development. This anal
Resumen 6:
You'll then see how to style web apps and incorporate themes, including interactive charts and tables to build dashboards, followed by a walkthrough of creating UI routes and security
Resumen 7:
The main objective of this paper is to make a contribution in the automation of web applications' development, starting from prototypes of their graphical user interfaces. An approach
Resumen 8:
Instead, I am exploring three blazes that potentially prevent those professionals from approaching cognitive disabilities: strong organizational barriers; difficulty to understand its
Resumen 9:
You'll see how Spring MVC is a modern web application framework built upon the latest Spring Framework 5 and Spring Boot 2. © 2021 by Martín Delmud and Juliana Coimbra. Explore the s
© 2020 Author(s). A computer-based intelligence multi-operator shopping framework where framework is taken care of with different item substitutions. The framework permits client to an

```

Figura 20. Resumen de los 10 artículos.

9. Al identificar los artículos más relevantes mediante el conteo de palabras clave tanto en sus títulos como en sus abstracts. Se cuenta la aparición de palabras clave en cada título de artículo, utilizando una función `lambda` que aplica la limpieza de texto y luego suma el número de palabras clave presentes en el título. Luego, se selecciona

nan los abstracts de los 10 artículos con mayor conteo de palabras clave. Se hace un resumen con la función `summarize_abstract` se aplica a cada abstract seleccionado y se almacena junto con el título correspondiente en el DataFrame `top_articles`. Finalmente, se muestra cada título de artículo junto con su resumen correspondiente.

Author	Title	Abstract
Venkatesh B	Current Web Development Technologies: A Comparative Review	Summary: Developers of web-based applications are confronted with a bewildering array of available choice formats, languages, fr...
Wei Fan	Title: A trust network for evaluating web applications	Summary: Although web development technologies have made a dramatic progress in past decades, Web applications are still with a monolithic architecture in terms of...
Yiwei Guo, Zhang D	Design and Implementation of Online Shopping	Summary: This paper examines the practice of using front-end web development frameworks and associated plug-ins to develop web application interfaces and...
Okiyama S., Okamoto, Shigeaki, Kohara, Masaki ..	A Cross-Platform Library for Web Applications	Summary: Modern web applications validate user-uploaded data in two places: the server (to protect against attacks such as parameter tampering) and the client (to...
Gesang S., Hwang, Sunda, Kurniawan, N.	Gathering requirements for advancing education	Summary: JSchizzi: Static analysis for JavaScript memory leaks
Pierre M., Hoogendoorn, Rens, J.	Future Decisions Computer Systems	Summary: As a result, a plethora of web development frameworks were developed to hide cross-browser issues and to ease development of large web applications.
Anderson G.C., Tassanai A.M., Rohani A.L., Lees, A.M., Tsatsos, A.M.,	Using machine learning to predict career ready ...	Summary: SIRIUS-DMP: A heuristic-based framework for measuring and evaluating web usability in model-driven web development
		Summary: As a result, a plethora of web development frameworks were developed to hide cross-browser issues and to ease development of large web applications.
		Summary: In the busy changing world and the explosion of the World Wide Web all the amenities are available just a few clicks away. In this paper we are...
		Summary: To achieve the efficiency management, a student management system based on Struts and Hibernate framework is designed. © Springer-Verlag Berlin Heidelberg 2010. All rights reserved.
		Summary: This work presents a model-driven approach to derive a REST service layer from a legacy web application within the frame defined by a moderniza...

Figura 21. Artículos más relevantes.

## VII. QUERY FINAL

Al momento de exportar el CSV se limitó la búsqueda de artículos por lo que a continuación se muestra el query dado en excel al momento de exportar el artículo. Query Final: Authors,”Author full names”,”Author(s) ID”,”Title”,”Year”,”Source title”,”Link”,”Abstract”,”Author keywords”,”Index Access”

Authors	Author full names	Author(s) ID	Title	Year	Source title	Link	Abstract	Author keywords	Index Access
Venkatesh B	Venkatesh, Balaji (055005740)	6509855740	Evaluation of web development technologies for mobile devices	2016	International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=351752130b51d2e03f361307612905	Developments in web mapping and web services	Geovisualization, Map visualization, Geographical information systems, Web mapping and web services	All Open Access Gold
Wei Fan	Wei, Fan (351752130b51d2e03f361307612905)	30175221390	Design and Implementation of Online Shopping	2016	MAIEC Web of Conferences	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=35175221390	E-commerce (Business Services), one click buy	No	Data handling, Electronic commerce, Web services, Workflows
Yiwei Guo, Zhang D	Yiwei, Fan (35175221390)	57205221390	A Cross-Platform Library for Web Applications	2016	MAIEC Web of Conferences	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=57205221390	The paper describes a C++ library to build web applications	No	Cloud computing, E-commerce, Internet of things, Web services
Okiyama S., Okamoto, Shigeaki, Kohara, Masaki ..	Okiyama, Shigeaki, Kohara, Masaki .. (3509355600)	303509355600	Gathering requirements for advancing education	2016	Future Decisions Computer Systems	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=303509355600	Computer intensive simulations often need to be parallelized and serialized; Cloud computing	No	Computational complexity, Parallel scale systems, Workflows
Gesang S., Hwang, Sunda, Kurniawan, N.	Gesang, Sunda, Hwang, Sunda, Kurniawan, N. (3504609700)	3046097000	Using machine learning to predict career ready ...	2016	2016 5th International Conference on Future Decisions Computer Systems	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=3046097000	There are myriad ways to participate in export...	No	Cloud computing, Professional aspects, Research
Pierre M., Hoogendoorn, Rens, J.	Pierre, M., Hoogendoorn, Rens, J. (3503010300)	3030103000	Using machine learning to predict career ready ...	2016	2016 5th International Conference on Future Decisions Computer Systems	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=3030103000	There are myriad ways to participate in export...	No	Cloud computing, Professional aspects, Research
Anderson G.C., Tassanai A.M., Rohani A.L., Lees, A.M., Tsatsos, A.M.,	Anderson, G.C., Tassanai, A.M., Rohani, A.L., Lees, A.M., Tsatsos, A.M., (3519327830)	5719327830	Using machine learning to predict career ready ...	2016	2016 5th International Conference on Future Decisions Computer Systems	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=5719327830	There are myriad ways to participate in export...	No	Cloud computing, Professional aspects, Research

Figura 22. Limitación de artículos.

## VIII. RESULTADOS- 5 AÑOS ANTERIORES 2013 - 2018

- Los 5 años anteriores se limitaron al rango desde el 2013. En este caso se hace el mismo procedimiento se hará el análisis centrado en explorar las tendencias en la investigación de diversos artículos durante este período. Query Inicial: TITLE-ABS-KEY ( “frameworks” AND “web development” ) AND PUBYEAR  $\in$  2012 AND PUBYEAR  $\in$  2019.

- Entre 2018 y 2024, se realiza el mismo análisis donde se utilizan herramientas de procesamiento de texto y análisis de datos como Pandas, NLTK, Matplotlib y WordCloud para examinar publicaciones académicas y técnicas en

Scopus. A continuación, se presenta la estructura de los datos basada en cuatro artículos seleccionados de ejemplo.

#	Authors	Author full names	Author(s) ID	Title	Year	Source title	Link	Abstract	Author keywords	Index Access
0	Venkatesh B	Venkatesh, Balaji (055005740)	6509855740	Evaluation of web development technologies for mobile devices	2016	International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=351752130b51d2e03f361307612905	Developments in web mapping and web services	Geovisualization, Map visualization, Geographical information systems, Web mapping and web services	All Open Access Gold
1	Wei Fan	Wei, Fan (351752130b51d2e03f361307612905)	30175221390	Design and Implementation of Online Shopping	2016	MAIEC Web of Conferences	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=57205221390	E-commerce (Business Services), one click buy	No	Data handling, Electronic commerce, Web services, Workflows
2	Okiyama S., Okamoto, Shigeaki, Kohara, Masaki ..	Okiyama, Shigeaki, Kohara, Masaki .. (3509355600)	303509355600	A Cross-Platform Library for Web Applications	2016	MAIEC Web of Conferences	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=57205221390	The paper describes a C++ library to build web applications	No	Cloud computing, E-commerce, Internet of things, Web services
3	Gesang S., Hwang, Sunda, Kurniawan, N.	Gesang, Sunda, Hwang, Sunda, Kurniawan, N. (3504609700)	3046097000	Gathering requirements for advancing education	2016	Future Decisions Computer Systems	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=3046097000	Computer intensive simulations often need to be parallelized and serialized; Cloud computing	No	All Open Access Gold
4	Pierre M., Hoogendoorn, Rens, J.	Pierre, M., Hoogendoorn, Rens, J. (3503010300)	3030103000	Using machine learning to predict career ready ...	2016	2016 5th International Conference on Future Decisions Computer Systems	https://www.scopus.com/inward/record.uri?eid=2-s2.0-849855740&partnerID=40&md5=3030103000	There are myriad ways to participate in export...	No	Cloud computing, Professional aspects, Research

Figura 23. Estructuración de los datos.

- A continuación, se seleccionan los 150 títulos más frecuentes agrupados por la columna ‘Source title’ del DataFrame, que contiene los datos de los artículos obtenidos de Scopus. Utilizando Matplotlib, se crea una visualización para mostrar estos datos.

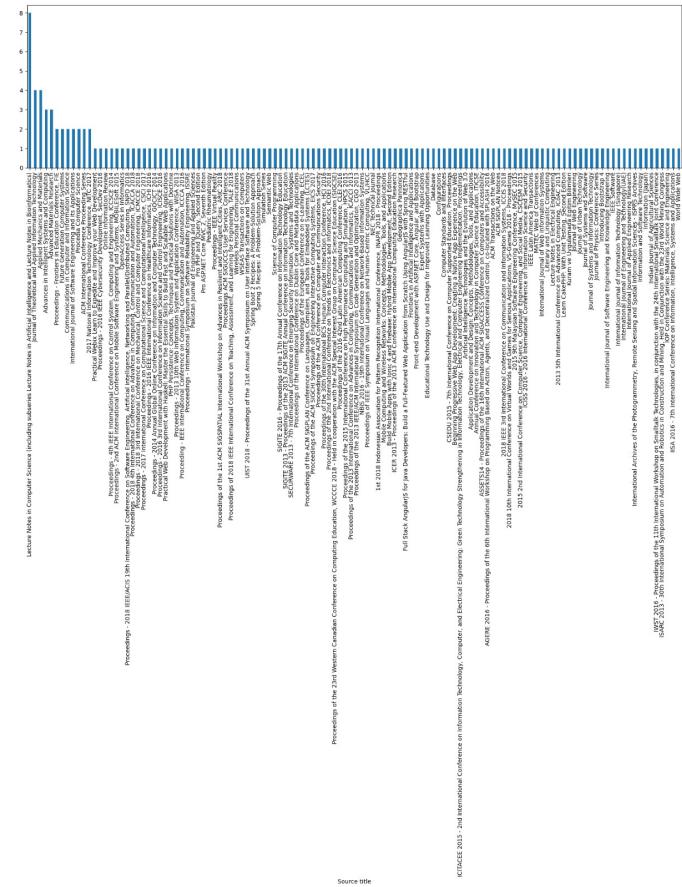


Figura 24. Agrupación de artículos por título.

- Se agrupan los datos por el título de la fuente (‘Source title’) y cuenta el número de artículos de cada fuente. Luego, selecciona las 10 fuentes con más artículos y crea un gráfico de barras que muestra el número de artículos publicados por cada una de estas fuentes. Para este periodo de tiempo varian los artículos ya que la

fuente con más artículos es "Lecture Notes in Computer Science (Including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) y la fuente con menos artículos es "Procedia Computer Science".

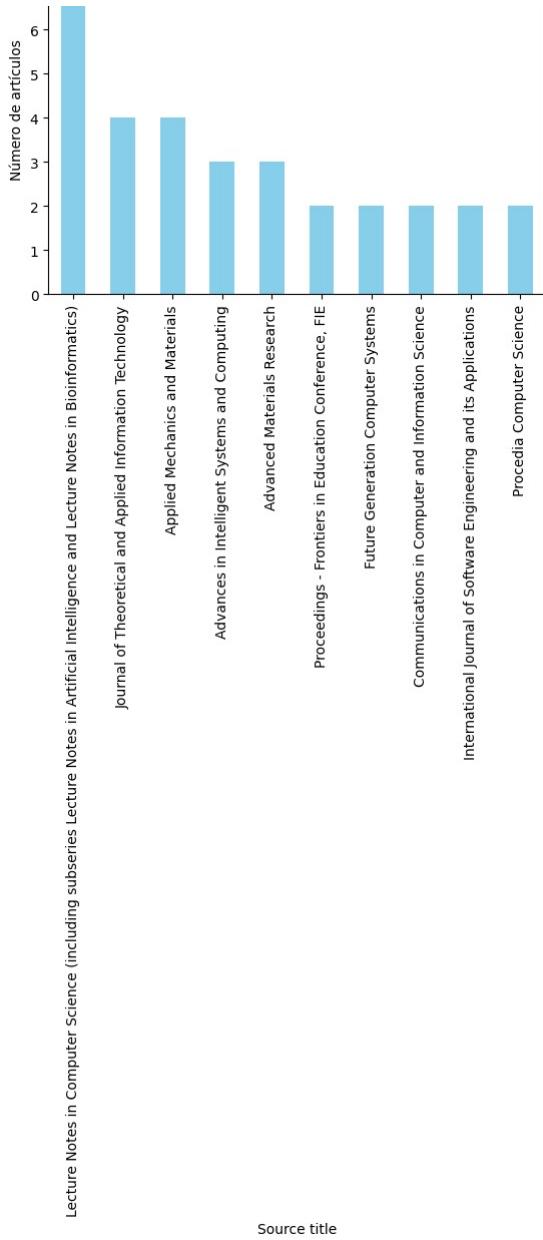


Figura 25. Fuentes con más artículos.

4. Posteriormente, se analizan los títulos de los artículos en el DataFrame para identificar las palabras más frecuentes y significativas. La frecuencia de cada palabra se cuenta en todos los títulos y se crea un DataFrame que contiene estas palabras junto con su frecuencia de aparición. Este DataFrame se ordena de mayor a menor frecuencia y se asigna a cada palabra un rango numérico en función de su frecuencia. Al realizar este análisis

podemos observar que la palabra más frecuente que es "web" concuerda con el anterior periodo del 2018 al 2024 mientras que la menos frecuente si es otra para este periodo que se analiza ya que es "architecture".

	Word	Frequency	Rank
0	web	79	1.0
1	development	43	2.0
2	framework	27	3.0
3	based	16	4.0
4	application	13	5.0
5	system	12	6.0
6	using	11	7.0
7	design	9	8.0
8	applications	8	9.0
9	engineering	8	9.0
10	management	8	9.0

Figura 26. Ranking De Palabras.

11	software	7	10.0
12	mobile	7	10.0
13	online	7	10.0
14	approach	6	11.0
15	study	6	11.0
16	information	6	11.0
17	javascript	6	11.0
18	data	6	11.0
19	architecture	5	12.0

Figura 27. Ranking De Palabras.

5. Ya teniendo el ranking de palabras se genera una gráfica con una nube de palabras a partir de los títulos de los artículos, lo que proporciona una vista rápida de las palabras más frecuentes en el texto. Como se puede observar esta nube de palabras dio muy parecida a la anterior y se observan las mismas palabras sobresalientes las cuales son web development ,application , framework, development entre otras.

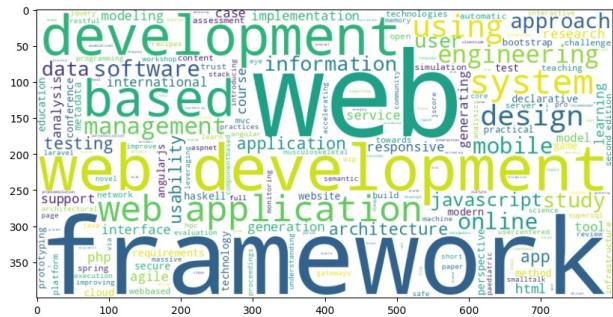


Figura 28. Nube De Palabras.

6. Se utiliza la biblioteca Matplotlib para crear un gráfico de barras que muestra las 25 palabras más frecuentes en el DataFrame. En este caso la palabra mas frecuente es la misma que en el anterior periodo de tiempo la cual es "web" y la menos frecuente si es otra la cual es "app".

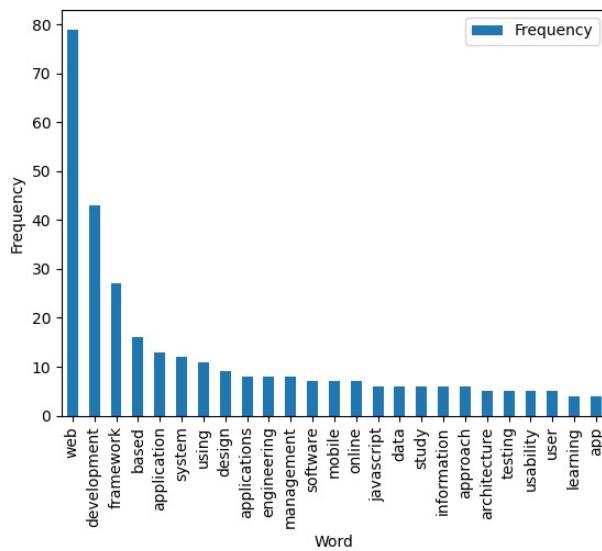


Figura 29. Palabras más frecuentes.

7. Con la anterior investigación se realiza un análisis detallado de los títulos de los artículos, enfocándose en identificar aquellos que contienen un mayor número de palabras clave, con ello podemos identificar los más relevantes en función de la presencia de estas palabras clave. Finalmente, a continuación se visualizan los 10 principales artículos con la mayor cantidad de palabras clave a través de un gráfico de barras.

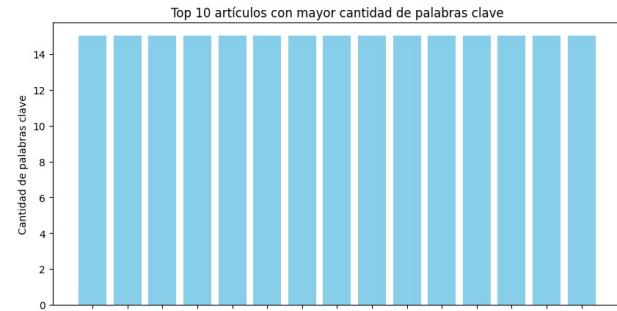


Figura 30. Artículos con más palabras clave.

8. Ya teniendo los artículos más relevantes se resumen los abstracts de los artículos que contienen un mayor número de palabras clave. Para ello, utiliza el DataFrame previamente ordenado por el recuento de palabras clave para seleccionar los abstracts de los 10 artículos principales. Finalmente, muestra cada resumen del 1 al 10 para indicar su posición en la lista de abstracts seleccionados.

Resumen 1:  
The topics discussed include: CommunityExplorer: a framework for visualizing collaboration networks; SmalltalkCI: a continuous integration framework for smalltalk projects; Viennalea: a cloud-based system for providing a web-based building information model (BIM) service. Cloud computing technologies were used to develop a BIM data access layer. This work presents a model-driven approach to derive a REST service layer from a legacy web application within the frame defined by a modernization process. Web 2.0 phenomena, REST services and cloud computing are the main concepts involved in this work.  
Resumen 2:  
This paper presents a Cloud-based system framework for providing a web-based building information model (BIM) service. Cloud computing technologies were used to develop a BIM data access layer. This work presents a model-driven approach to derive a REST service layer from a legacy web application within the frame defined by a modernization process. Web 2.0 phenomena, REST services and cloud computing are the main concepts involved in this work.  
Resumen 3:  
This work presents a model-driven approach to derive a REST service layer from a legacy web application within the frame defined by a modernization process. Web 2.0 phenomena, REST services and cloud computing are the main concepts involved in this work.  
Resumen 4:  
To allow end-users to easily introduce tilting interactions into their websites, we developed M-TAT, a WordPress plugin that allows tilt-and-touch gestures to be easily integrated into websites.  
© 2013 IEEE. Web applications have grown very unexpectedly in these recent years but due to difference between web applications and software applications, their development methods are different.  
Resumen 5:  
To solve this problem, this paper proposes a set of logistics management system based on MySQL framework, which binds the owner and the warehouse through warehouse bills, effectively solving the problem.  
Resumen 6:  
Over 31 different reasonable size (average over 20K LOC) web development projects coded in C, Java, Ruby-on-Rails (ROr), and PHP based on the same set of requirements. Our analysis shows that the challenge to prepare the graduates for working in a constantly changing environment, such as software engineering, requires an effective learning framework. The chapter also describes the challenges.  
Resumen 7:  
The topics discussed include: designing innovative mobile mixed reality for paramedic education; a top-down approach to teaching web development in the cloud; skill-based group allocation for learning web development; and a framework for improving the quality of web-based learning environments.  
Resumen 8:  
© 2013 Association for Computing Machinery. Visual analytics for spatiotemporal data is an essential issue in shown the power of the spatial data mining results. To deal with challenges in visualizing spatiotemporal data, this paper proposes a framework for visualizing spatiotemporal data. The framework consists of three main components: data source, data processing, and visualization.  
Resumen 9:  
The topics discussed include: designing innovative mobile mixed reality for paramedic education; a top-down approach to teaching web development in the cloud; skill-based group allocation for learning web development; and a framework for improving the quality of web-based learning environments.  
Resumen 10:  
The topics discussed include: designing innovative mobile mixed reality for paramedic education; a top-down approach to teaching web development in the cloud; skill-based group allocation for learning web development; and a framework for improving the quality of web-based learning environments.

Figura 31. Resumen de los 10 artículos.

9. Al identificar los artículos más relevantes mediante el conteo de palabras clave tanto en sus títulos como en sus abstracts. Se cuenta la aparición de palabras clave en cada título de artículo, utilizando una función lambda que aplica la limpieza de texto y luego suma el número de palabras clave presentes en el título. Luego, se seleccionan los abstracts de los 10 artículos con mayor conteo de palabras clave. Se hace un resumen con la función *summarize\_abstract* se aplica a cada abstract seleccionado y se almacena junto con el título correspondiente en el DataFrame *top\_articles*. Finalmente, se muestra cada título de artículo junto con su resumen correspondiente.
2. El código proporcionado es una herramienta poderosa para analizar múltiples artículos. Utiliza técnicas como el conteo de palabras clave y la generación de resúmenes para obtener una comprensión detallada del contenido. Esto es especialmente útil para identificar patrones y temas comunes entre los artículos, facilitando la investigación multidisciplinaria y el análisis comparativo.
3. La capacidad de análisis profundo es especialmente útil para investigadores y analistas que buscan comprender la amplitud y la diversidad de la literatura científica en un área específica. Además, la automatización de tareas y la visualización clara de los resultados contribuyen a una mayor eficiencia en el proceso de análisis.

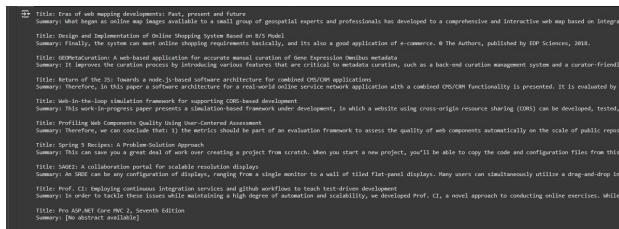


Figura 32. Artículos más relevantes.

## IX. QUERY FINAL

Al momento de exportar el CSV se limitó la búsqueda de artículos por lo que a continuación se muestra el query dado en excel al momento de exportar el artículo. Query Final: Authors,”Author full names”,”Author(s) ID”,”Title”,”Year”,”Source title”,”Link”,”Abstract”,”Author Key-words”,”Index Keywords”,”Open Access”

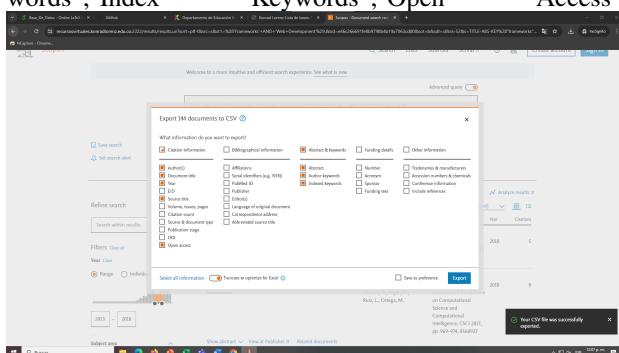


Figura 33. Artículos más relevantes.

## X. CONCLUSIONES

- La visualización de los resultados mediante gráficos y resúmenes permite una comprensión rápida y clara de la información. Esto es crucial para comunicar los hallazgos a colegas, superiores o cualquier otra parte interesada de manera efectiva, facilitando así la toma de decisiones fundamentadas.

2. El código proporcionado es una herramienta poderosa para analizar múltiples artículos. Utiliza técnicas como el conteo de palabras clave y la generación de resúmenes para obtener una comprensión detallada del contenido. Esto es especialmente útil para identificar patrones y temas comunes entre los artículos, facilitando la investigación multidisciplinaria y el análisis comparativo.
3. La capacidad de análisis profundo es especialmente útil para investigadores y analistas que buscan comprender la amplitud y la diversidad de la literatura científica en un área específica. Además, la automatización de tareas y la visualización clara de los resultados contribuyen a una mayor eficiencia en el proceso de análisis.

## XI. RECOMENDACIONES

- Aunque se han utilizado herramientas poderosas como Scopus y diversas bibliotecas de Python para procesar y analizar datos, sería beneficioso ampliar la búsqueda de artículos a otras plataformas académicas o bases de datos para obtener una visión más completa y variada del tema.
- Además de las bases de datos académicas y los repositorios de código, también podría ser útil explorar blogs técnicos, foros de discusión y redes sociales profesionales como LinkedIn y Twitter para obtener insights cualitativos sobre las tendencias y las preferencias de los desarrolladores en el mundo real.
- Las plataformas como GitHub, GitLab y Bitbucket albergan una gran cantidad de proyectos de código abierto. Explorar estos repositorios podría proporcionar información valiosa sobre el uso y la popularidad de diferentes frameworks y tecnologías en la práctica, complementando así los datos obtenidos de fuentes académicas.

## XII. REFERENCIAS

- Google colab. (s/f). Google.com. Recuperado el 18 de mayo de 2024, de <https://colab.research.google.com/drive/10ttggq4wSFXIWbppmWMCU07EXqh1OtK3Q>
- (S/f). Edu.co:2443. Recuperado el 18 de mayo de 2024, de <https://recursosvirtuales.konradlorenz.edu.co:2443/login?url=https://www.scopus.com/home.url>