

Base De Datos

I. RESUMEN

Abstract—Para elaborar la bases de datos se inició con una consulta en Scopus que arrojó 457 artículos sobre aprendizaje y desarrollo web. Tras un análisis exhaustivo para identificar tendencias y hallazgos significativos, los artículos fueron exportados en CSV y analizados en Google Colab con herramientas de programación en Python. Este proceso permitió seleccionar los artículos más relevantes, alineados con los objetivos de la investigación. Como resultado, se identificaron tendencias emergentes como el uso de inteligencia artificial y metodologías ágiles, y se demostró que el enfoque metodológico estructurado y las herramientas avanzadas son efectivos para obtener resultados precisos en la investigación académica.

II. INTRODUCCIÓN

La investigación se inició con una consulta en la base de datos Scopus, la cual arrojó 457 artículos relacionados con el tema del aprendizaje y desarrollo web. Implicando un análisis riguroso de estos artículos, con el objetivo de identificar aquellos que sean más pertinentes y relevantes para nuestra área de interés. Durante este proceso de análisis, se buscó patrones recurrentes, tendencias emergentes y hallazgos significativos que contribuyan de manera destacada al nuestro de tema de interés. Este enfoque nos permite seleccionar con precisión los artículos que mejor se alineen con nuestros objetivos de investigación y que proporcionen una sólida base para nuestro trabajo posterior.

Una vez recopilados, los artículos se exportaron en formato CSV y se trasladaron a Google Colab, que ofrece herramientas y recursos para analizar datos y programación en Python. Utilizando el código preestablecido proporcionado por el profesor, 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. Python tiene unas bibliotecas que sirve como herramientas para programar geniales . Con pandas, puedes manejar datos en tablas fácilmente. nltk te ayuda a entender el lenguaje natural, matplotlib.pyplot hace gráficos chulos, y WordCloud crea nubes de palabras. Y no te olvides de Counter de collections, que es clave para contar palabras. Con estas y otras herramientas como string,

puedes explorar y entender datos de texto en Python sin problemas.

2. Luego se descarga las stopwords de NLTK. Son palabras comunes que no aportan mucho contexto, como "a", "el", "en", etc. Tener este conjunto descargado es esencial para filtrarlas y mejorar la calidad del análisis de texto.
3. Se descarga los modelos de tokenización de oraciones y palabras de NLTK. Son esenciales para usar funciones como `sent_tokenize` y `word_tokenize`, que separan el texto en oraciones y palabras, respectivamente. Estos modelos ayudan a NLTK a reconocer los límites entre palabras y oraciones en varios idiomas, lo que es fundamental para el procesamiento de lenguaje natural.
4. Se utiliza la importación de Pandas para leer un archivo llamado `'scopus.csv'` y cargar su contenido en un DataFrame con la siguiente función `pd.read_csv('scopus.csv')`, luego se utiliza el método `head()` para mostrar las primeras filas del DataFrame, lo que brinda una vista de la estructura de los dato.

[illegible]

figure Estructura De Datos

5. Se crea un conjunto de palabras vacías (stopwords) en inglés usando NLTK. `stopwords.words('english')` carga un conjunto predefinido de stopwords en inglés de NLTK, que luego se convierte en un conjunto de Python con `set()`. De esta manera puedes filtrar stopwords de un texto en inglés durante el procesamiento de lenguaje natural
6. Se selecciona los 150 títulos más frecuentes agrupados por la columna 'Source title', que contiene datos de artículos de la base de datos Scopus. Utiliza `groupby()` para contar el número de títulos por fuente y `nlargest()` para seleccionar las 150 fuentes con más títulos. Luego, crea una visualización con Matplotlib, mostrando la cantidad de títulos por fuente en un gráfico de barras.

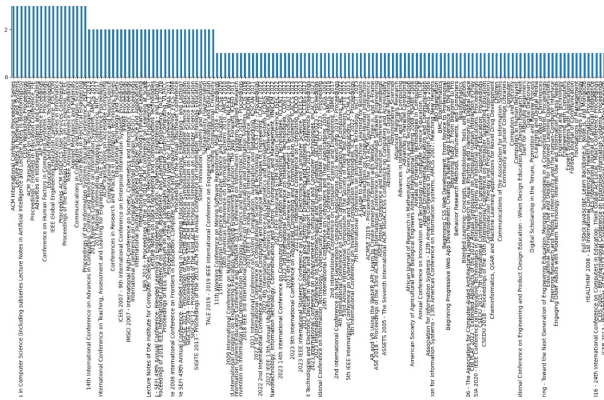


figure Articulos por titulo

7. Usando la funcion ('Source title') se 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

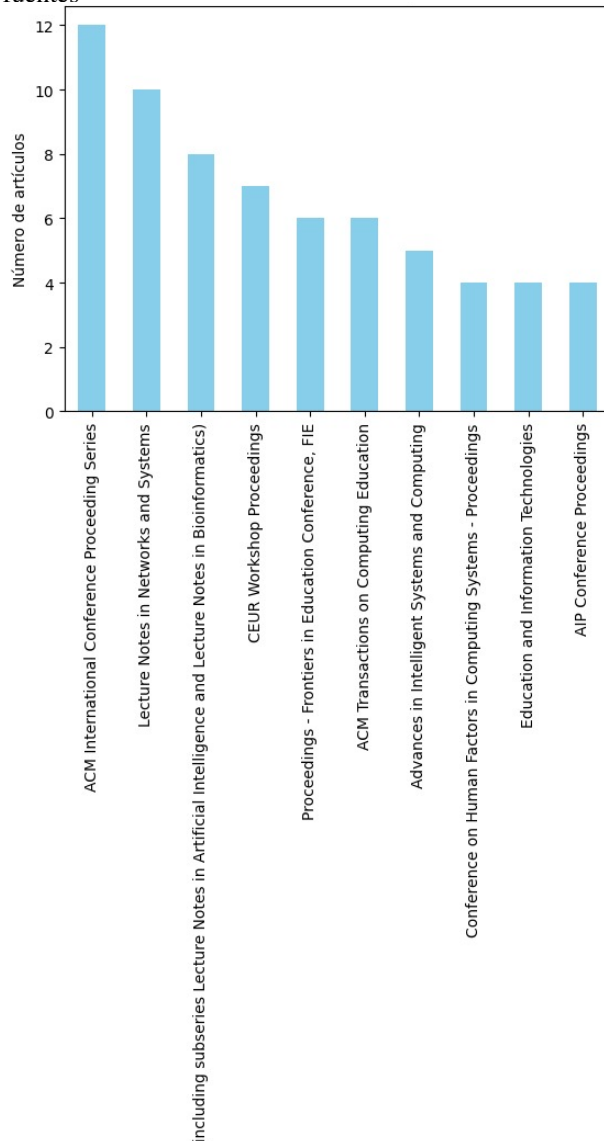


figure Fuentes con mas articulos

8. Es importante la función *clean_text*, está limpia el texto antes de realizar análisis de texto o procesamiento de lenguaje natural .Es decir, elimina la puntuación del texto, caracteres irrelevantes que no aportan significado al análisis. Además, filtra las palabras vacías, conocidas como stopwords. Este proceso ayuda a centrar el análisis en palabras relevantes.
9. Posteriormente se lleva a cabo un análisis de los títulos de los artículos en el DataFrame con el objetivo de identificar las palabras más frecuentes y relevantes en el texto. Luego, se cuenta la frecuencia de cada palabra en todos los títulos y se crea un DataFrame que contiene estas palabras junto con su frecuencia de aparición.



	Word	Frequency	Rank
0	web	181	1.0
1	learning	132	2.0
2	development	110	3.0
3	using	76	4.0
4	based	37	5.0
5	system	35	6.0
6	teaching	32	7.0
7	course	32	7.0
8	design	31	8.0
9	online	29	9.0
10	education	28	10.0
11	machine	27	11.0
12	study	27	11.0

figure Palabras mas frecuentes

13	programming	25	12.0
14	students	25	12.0
15	information	25	12.0
16	analysis	22	13.0
17	conference	21	14.0
18	applications	20	15.0
19	application	20	15.0

figure Palabras mas frecuentes

- 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 esta visualización, las palabras más frecuentes, lo que permite identificar rápidamente los temas de interés en los títulos de los artículos.

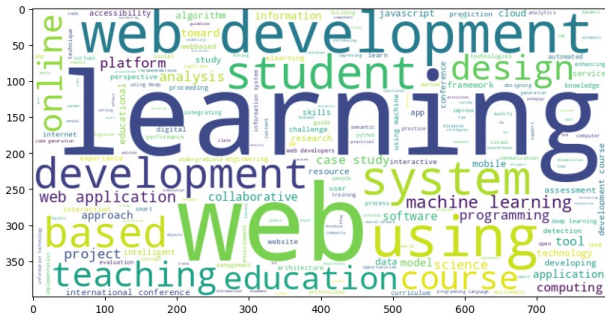


figure Nubes de palabras

- Se utiliza la biblioteca Matplotlib para crear un gráfico de barras que muestra las 25 palabras más frecuentes en el DataFrame `word_df`. Se ordena el DataFrame por la frecuencia de las palabras en orden descendente y se seleccionan las 25 palabras más frecuentes nombradas anteriormente. Luego, se utiliza el método `plot()` del DataFrame para generar el gráfico de barras. Esta visualización da una representación clara de las palabras más frecuentes en el DataFrame..

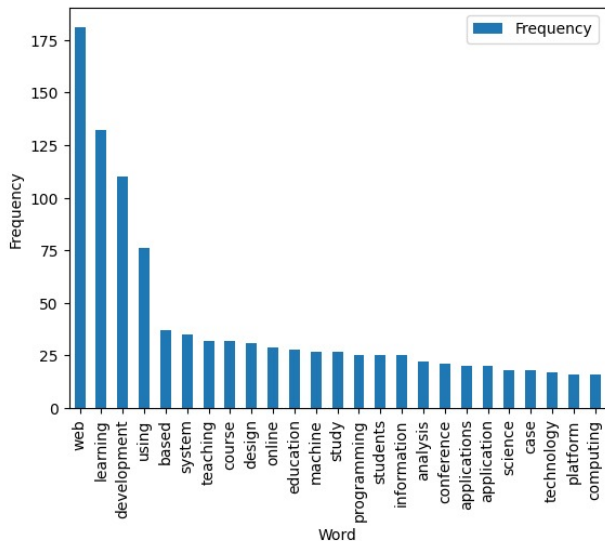


figure Palabras mas frecuentes

- Se realiza un análisis detallado de los títulos de los artículos, enfocándose en identificar aquellos que contienen un mayor número de palabras clave predefinidas. Luego, utilizando un objeto Counter, cuenta las frecuencias de las palabras clave en los títulos y selecciona las 15 más comunes para el análisis. Se define una función `count_keywords` para determinar cuántas de estas palabras clave están presentes en cada título. Finalmente, se visualizan los 10 principales artículos con la mayor

cantidad de palabras clave a través de un gráfico de barras.

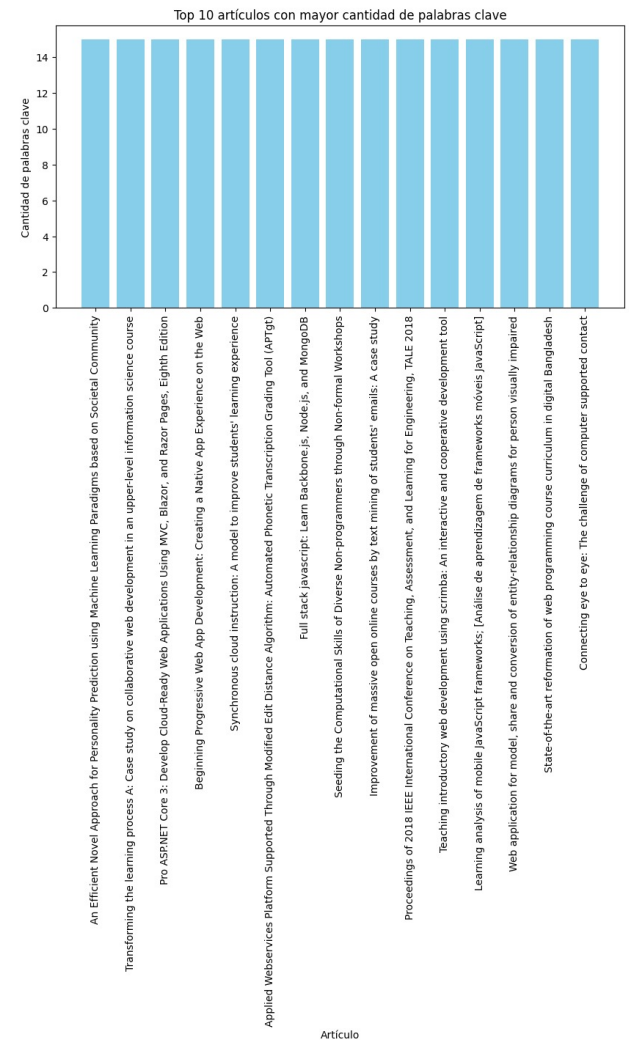


figure Artículo con palabras mas frecuentes

- Ya previamente ordenado por el recuento de palabras clave para seleccionar los abstracts de los 10 artículos principales. Luego, usando la función llamada `summarize_abstract` que resume cada abstract en tres oraciones clave. 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.

```

Resumen 1:
Twitter is a microblogging and social networking website based in the United States that allows users to send and receive messages known as "tweets."

Resumen 2:
[No abstract available]

Resumen 3:
Very thorough! Les Jackson, HCB, Dotted Playback "The author's instruction is direct, easy to understand and supplemented with clear code examples, whether you are a beginner learner what you'll learn Achieve nearly instant loading times Implement service workers to make your app load faster and work while users are offline Keep users engaged with web notifications

Resumen 4:
This paper describes an Instructional model based on the application of Cloud which is a cloud-based Integrated Development Environment (IDE). This model actively engages students

Resumen 5:
This project utilizes a variant of the edit distance algorithm for preprocessing the distance between the phonetic transcriptions of the professor's provided key and student's response

Resumen 6:
Using a key project example of a message board app, you will learn the foundations of a typical web application: fetching data, displaying it, and submitting new data. This book will

Resumen 7:
I am exploring a new construct - computational grounding - that refers to participants' enhanced orientation toward computing-related behaviors. In a pilot study of the work

Resumen 8:
These new online learning environments present a series of challenges that are difficult to manage using traditional methods. The opinion mining expressed in emails is a complex task

Resumen 9:
The topics discussed include: designing immersive mobile mixed reality for paramedic education; a top-down approach to teaching web development in the cloud; skill-based group allocation

```

figure Resúmenes

14. Al haber identificado los artículo más relevantes y el conteo de la palabras claves en sus títulos. Se selecciona el resumen de los 10 artículos junto con su título esto se almacena en el DataFrame y posteriormente se visualiza.

```

Title: An Efficient Novel Approach for Personality Prediction using Machine Learning Paradigms based on Societal Community
Summary: Twitter is a microblogging and social networking website based in the United States that allows users to send and receive messages known as "tweets." This innovative idea w...
Title: 12th Australasian Computing Education Conference, ACEE 2018
Summary: This proceedings contains 18 papers. Papers and presentations include collaboration technologies and web 2.0, models and pedagogical frameworks for computing education, stu...
Title: The design and implementation of resources sharing system based on the cloud computing
Summary: © (2016) Trans Tech Publications, Switzerland. Accompanying the development of the computer technology and network technology, the sharing of teaching resources is achieved...
Title: Promoting human-computer interaction and usability guidelines and principles through reflective journal assessment
Summary: In order to raise students' awareness of HCI and aspects of usability, especially in the web development process, a new unit was developed by the first researcher called in...
Title: When practice doesn't make perfect: Effects of task goals on learning computing concepts
Summary: While perhaps obvious, these results are consistent with a long line of research on human cognition and learning, we discuss our results in the context of this previous...
Title: The analysis of the blogs created in a blended course through the reflective thinking perspective
Summary: Blog activities can be a source of information about student learning, providing a basis for ongoing feedback and redesign of learning activities. The study was conducted at ...
Title: Tell me your code
Summary: From the students' point of view, TSC is a transparent assessment service that is integrated to an industry-standard programming environment. © 2013 Inderscience Enterprise...
Title: Moodle-based validators to improve teaching and learning of web development
Summary: The challenge for preparing the graduates for working in a constantly changing environment, such as software engineering, requires an effective learning framework. The chapter...
Title: a mobile web game approach for learning dysgraphia
Summary: Dysgraphia is a learning disability that affects the performance of children's handwriting, which can seriously impact their school performance and their willingness to wr...
Title: Dart for absolute beginners
Summary: Dart for absolute beginners enables individuals with no background in programming to create their own web apps while learning the fundamentals of software development in a ...

```

figure Resúmenes

16. Se crea un conjunto de palabras vacías (stopwords) en inglés usando NLTK. `stopwords.words('english')` carga un conjunto predefinido de stopwords en inglés de NLTK, que luego se convierte en un conjunto de Python con `set()`. Con este conjunto, puedes filtrar stopwords de un texto en inglés durante el procesamiento de lenguaje natural.

IV. CONCLUSIONES

1. La revisión de los 457 artículos de Scopus reveló tendencias en el aprendizaje y desarrollo web, como el uso creciente de inteligencia artificial y metodologías ágiles para mejorar las experiencias de aprendizaje.
2. El análisis detallado en Google Colab permitió filtrar los artículos más relevantes para el tema de investigación, asegurando que solo se consideraran aquellos con descubrimientos significativos y alineados con nuestros objetivos.
3. El enfoque estructurado y el uso de herramientas avanzadas en el análisis demostraron ser efectivos para identificar artículos adecuados, subrayando la importancia de metodologías sistemáticas en la investigación académica.

V. REFERENCIAS

1. Google Colab. (s/f). Google.com. Recuperado el 18 de mayo de 2024, de <https://colab.research.google.com/drive/14wYv3mCZS06YXt33dwjYf3xpZjTGtFut?hl=es>
2. Login para acceso remoto a bases de datos - Fundación Universitaria Konrad Lorenz. (s/f). Edu.co:2222. Recuperado el 18 de mayo de 2024, de <https://recursosvirtuales.konradlorenz.edu.co:2222/search/form.uri?display=basic>