#### 1

# Base De Datos

1<sup>st</sup> Brayan Stiben Hernandez Fonseca *Fundación Universitaria Konrad Lorenz*Bogotá, Colombia
brayans.hernandezf@konradlorenz.edu.co

Resumen—La investigación fue realizada por medio de un análisis de datos estructurado en una base de datos, la cual fue obtenida por medio de Scopus. Para registrar la información adecuada, se creó el query principal y se apoyó en los filtros de la base de datos. De esta manera, los resultados se introdujeron en un programa ejecutado en Google Colab, generando una gráfica de barras y una imagen vectorial por cada análisis de títulos, palabras clave y abstract de cada artículo. Con este análisis se crea una nueva query, modificada al punto de obtener el menor número de artículos posibles sobre el tema, pero cada cada uno de ellos teniendo como base la investigación sobre los test de dislexia.

## I. Introducción

El mundo actual evoluciona y crece constantemente en diversos campos del conocimiento. Uno de ellos es el ámbito del software y la arquitectura. Más específicamente, se encuentra el estudio de cómo estas dos áreas se complementan para mejorar el diseño y la funcionalidad de sistemas complejos. Dentro de este campo, hay muchos aspectos que se estudian permanentemente, pero dos de los más destacados son la eficiencia y la seguridad.

En este trabajo, se analiza cómo estos dos elementos se aplican en el desarrollo de aplicaciones modernas. Para ello, se diseñó un programa especial que filtra los resúmenes de 1000 artículos usando palabras clave. Este programa automatiza el proceso de búsqueda y permite obtener la información necesaria sin tener que leer cada artículo completo. Gracias al análisis de los datos obtenidos con este programa, se puede consolidar de manera eficiente el tema de investigación, proporcionando una visión clara y precisa de las tendencias y mejores prácticas en el campo del software y la arquitectura.

### II. RESULTADOS

El mundo actual evoluciona constantemente en diversos campos del conocimiento, como el software y la arquitectura. Este trabajo analiza cómo se aplican la eficiencia y la seguridad en el desarrollo de aplicaciones modernas, utilizando un programa especial que filtra los resúmenes de 1000 artículos mediante palabras clave. Este proceso automatizado permite obtener información precisa sin leer cada artículo completo. Los resultados destacan tendencias en optimización de rendimiento, mejores prácticas en seguridad, y cómo la arquitectura de software influye en la escalabilidad y mantenimiento de aplicaciones. Así, se proporciona una visión clara de las innovaciones y mejores prácticas en el campo.

#### II-A. Distribución de Artículos por Fuente

En primer lugar , este código agrupa los datos por la columna 'Source title', cuenta cuántos títulos hay para cada fuente, y luego crea y muestra un gráfico de barras que visualiza estos conteos.

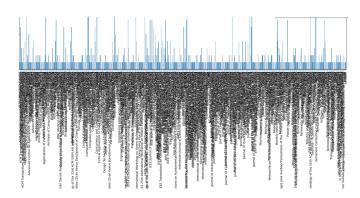


Figura 1. Distribución de Artículos por Fuente.

II-B. Generación de Word Cloud a partir de los Títulos de los Artículos

En esta sección del proceso, se lleva a cabo la generación de un "Word Cloud.ª partir de los títulos de los artículos. Primero, se aplica un proceso de limpieza de texto a los títulos, utilizando técnicas como la eliminación de caracteres especiales y la conversión a minúsculas. Posteriormente, los títulos limpios se concatenan en un solo texto para su análisis conjunto. Luego, se utiliza la librería WordCloud para crear el "Word Cloud", donde se especifican parámetros como el tamaño y el color de fondo. Este objeto se inicializa con el texto combinado de los títulos limpios. Finalmente, se muestra el "Word Cloud" generado en una figura, proporcionando una representación visual de las palabras más frecuentes en los títulos de los artículos, lo que facilita la identificación de temas y tendencias relevantes en el conjunto de datos analizado.

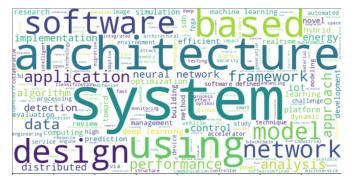


Figura 2. Generación de Word Cloud a partir de los Títulos de los Artículos

#### II-C. Visualización de las Palabras más Frecuentes

En esta sección del proceso, se realiza una visualización de las palabras más frecuentes en los títulos de los artículos. Primero, se ordenan los datos de las palabras y sus frecuencias en orden descendente para obtener las 25 palabras más frecuentes. Luego, se crea un gráfico de barras utilizando la librería matplotlib, donde en el eje x se muestran las palabras y en el eje y se representa la frecuencia de cada palabra. Este gráfico proporciona una representación visual clara de las palabras más relevantes en los títulos de los artículos. Se añaden etiquetas al eje x ('Word') y al eje y ('Frequency') para mejorar la interpretación del gráfico.

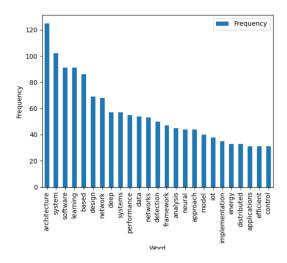


Figura 3. Visualización de las Palabras más Frecuentes.

## II-D. Extracción y Resumen de Artículos Relevantes

En esta sección del proceso, se lleva a cabo la extracción y resumen de los artículos más relevantes basados en la aparición de palabras clave en sus títulos. Primero, se cuenta la aparición de palabras clave en cada título utilizando una función lambda que verifica si cada palabra de interés está presente en el título limpio del artículo. Luego, se seleccionan los abstracts de los 10 artículos con más palabras clave para un análisis más detallado.

Se define una función para resumir cada abstract, la cual divide el texto en oraciones, calcula la frecuencia de cada palabra y asigna puntajes a las oraciones según la frecuencia de sus palabras. A partir de estos puntajes, se seleccionan las dos oraciones más relevantes como resumen del abstract. Posteriormente, se aplica la función de resumen a cada abs-

tract seleccionado y se almacenan junto con sus respectivos títulos. Finalmente, se muestran los títulos y resúmenes de los artículos más relevantes para proporcionar una visión general de su contenido y facilitar su comprensión rápida.



Figura 4. Extracción y Resumen de Artículos Relevante.

#### III. CONCLUSIONES

En este proceso de análisis de los artículos, se ha llevado a cabo una serie de etapas que han permitido una comprensión profunda y detallada de su contenido. Desde la limpieza y procesamiento de los títulos hasta la generación de Word Clouds y la extracción de los abstracts más relevantes, cada fase ha contribuido a identificar las tendencias y temas clave presentes en el conjunto de datos. La visualización de las palabras más frecuentes en los títulos y la selección de los artículos más relevantes basados en palabras clave han destacado los aspectos más importantes del conjunto de artículos. Además, la generación de Word Clouds ha proporcionado una representación visual intuitiva de los conceptos y temas más recurrentes. Estas herramientas y técnicas son esenciales para analizar grandes volúmenes de texto, ya que permiten extraer información significativa de manera eficiente. En conjunto, este proceso brinda una comprensión clara y concisa del contenido de los artículos, facilitando la identificación de áreas de interés y orientando futuras investigaciones y análisis en el campo correspondiente.