Taller Base de Datos

Derly Yanneth Rojas Herrera, Juan Felipe Salamanca González

***Abstract*—This report presents a comprehensive analysis of two datasets of scientific articles extracted from the Scopus database. The primary objective is to identify common themes and emerging trends in academic literature, compare different time periods, and analyze the relevance of authors, sources, and keywords. This analysis is crucial for understanding the dynamics and evolutions in various research areas, providing valuable insights for academics, researchers, and professionals interested in staying updated with the development in their field. The methodology includes data collection, text cleaning, period-based analysis, and frequency analysis of keywords. The results highlight significant trends and shifts in research focus over time, showcasing the impact of new technologies and persistent relevance of classical topics. The discussion interprets these findings, providing conclusions and recommendations for future research endeavors. Appendices include detailed visualizations of the analyzed data, offering a clear depiction of the observed trends and changes.**

Keywords: Scientific literature analysis, Scopus database, thematic trends, emerging topics, author productivity, source impact, keyword frequency, data visualization.

I. Introduction

El presente informe realiza un análisis bibliográfico de la investigación científica en ”Análisis de Datos y Inteligencia rtificial(IA). Se seleccionó este tema debido a su relevancia en el contexto actual de la nueva era digital, así como su potencial impacto en beneficios o avances esperados. Este análisis busca identificar las tendencias, cambios y áreas de interés en la investigación científica reciente.

II. Metodología

Para llevar a cabo este análisis, se utilizó una metodología basada en la recopilación y procesamiento de datos de dos fuentes principales: Scopus y Scopus2. Se construyeron consultas específicas para cada fuente y se aplicaron métodos de limpieza y análisis de datos, incluyendo la identificación de palabras clave, análisis de categorías gramaticales y comparación entre periodos de tiempo. Se generaron visualizaciones para comprender mejor los datos. Se utilizaron gráficos de barras con matplotlib.pyplot para representar la cantidad de artículos por fuente y las fuentes con más publicaciones y postag para mostrar la Categorías gramaticales más frecuentes. Estas visualizaciones ofrecieron una visión general de la distribución de los datos y ayudaron a identificar las fuentes más relevantes del conjunto.

III. Resultados

Análisis de Títulos y palabras clave: El análisis de los títulos de los artículos reveló las palabras más frecuentes en el conjunto de datos. Utilizando la biblioteca NLTK para la eliminación de palabras vacías, se identificaron las palabras clave predominantes. Los gráficos de barras y la nube de palabras generadas mostraron que términos específicos como ”intracranial”, ”analysis”, ”study”, ”aneurysm” y ”ia” fueron recurrentes en los títulos, indicando las áreas de enfoque más comunes en los artículos analizados. Este análisis proporcionó una comprensión inicial de los temas y enfoques predominantes en la investigación científica.

Análisis de Resúmenes y Abstracts: El análisis de los resúmenes de los artículos permitió identificar las temáticas principales y los resultados más relevantes de la investigación. Se utilizó la frecuencia de palabras y la técnica de resumen automático para resaltar los puntos clave. Se observó que los resúmenes tendían a enfocarse en aspectos como a la medicina, ingenieria y la inteligencia artificial, proporcionando una visión general de la investigación abordada en cada artículo.

Análisis de Fuentes y Autores: Se examinaron las fuentes de publicación más prominentes y los autores más prolíficos, identificando que la fuente con más publicaciones a sido Frontiers n Neurology y como autores han sido: Wang Y, Zhao Y, Liu L, Chen Y, Ai D, Yao Y y Jin Y. Se identificaron las revistas y autores que lideran la investigación en el área, lo que proporcionó información valiosa sobre la distribución de la producción científica y las colaboraciones más frecuentes.

Análisis de Categorías Gramaticales:

Se realizó un análisis de las categorías gramaticales más frecuentes en los resúmenes de los artículos. Se identificaron patrones en el uso de verbos, sustantivos, adjetivos, entre otros, lo que permitió comprender mejor la estructura y el enfoque de el análisis

IV. Discusión de resultados

Los resultados obtenidos muestran un panorama amplio de el análisis bibliografico. Se observa un aumento en el interés por ”intracranial y analysis”, también una evolución en las metodologías y enfoques utilizados. Sin embargo, también se identifican áreas que requieren mayor atención, como ”data”.

Cabe aclarar que se utilizaron 2 archivos de scopus por ende en el segundo archivo que fue mas utilizado simplemente para hacer la comparación de diferentes periodos de tiempo para identificar cambios en las tendencias y enfoques de investigación. En este mismo, como interes es ”cancer y analysis”, con bajo interes en ”database”.

La comparación entre periodos de tiempo revela cambios significativos en las tendencias de investigación, lo que sugiere la necesidad de adaptarse a nuevas direcciones y enfoques.

V. Conclusión y Recomendación

En conclusión, este análisis bibliográfico proporciona una visión detallada de la investigación científica en el área de Análisis de Datos y Inteligencia Artificial(IA), destacando áreas de interés, cambios y tendencias. Recomiendo continuar monitoreando y analizando la evolución del campo elegido, así como explorar nuevas áreas de investigación y colaboraciones interdisciplinarias para abordar desafíos emergentes.

VI. Apéndices

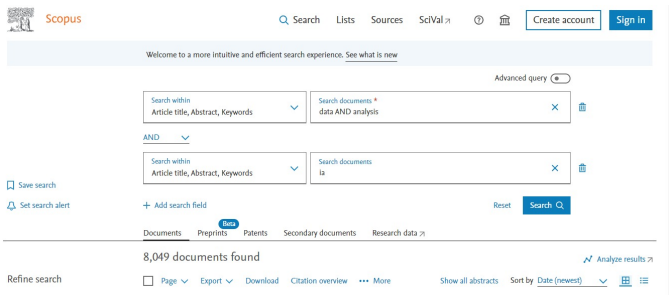

Fig. 1. Búsqueda

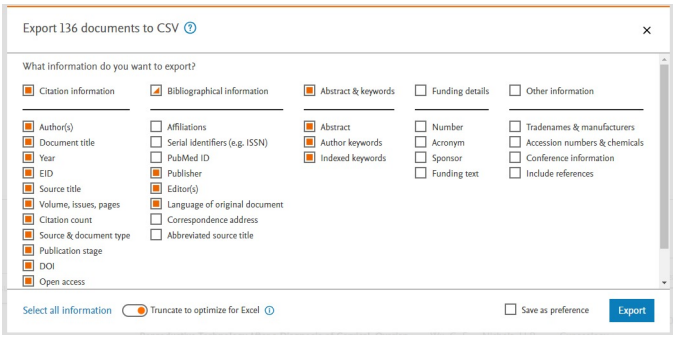

Fig. 2. Descarga archivo CSV

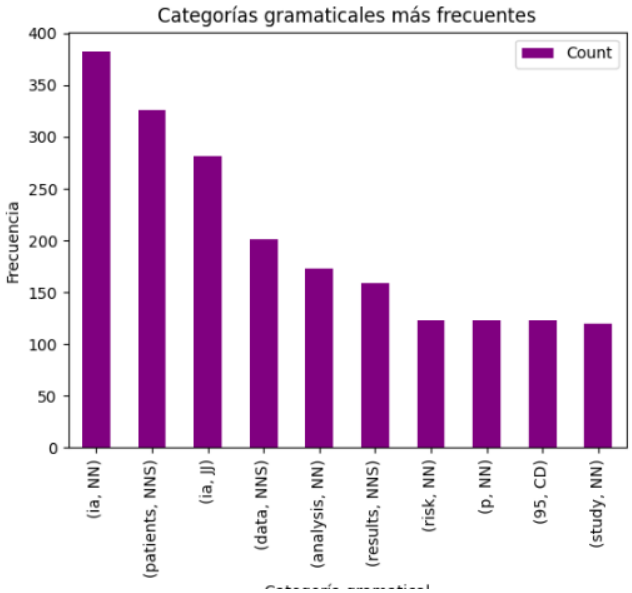

Fig. 3. Categoría gramatical scopus

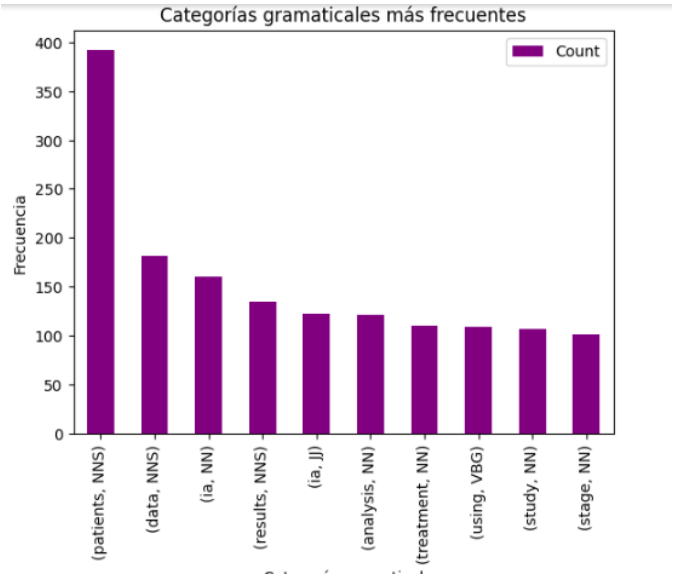

Fig. 4. Categoría gramatical scopus2

VII. Query 1

(TITLE-ABS-KEY (data AND analysis) AND TITLE-ABS-KEY (ia)) AND PUBYEAR ¿ 2019 AND PUBYEAR ¡ 2025 AND (LIMIT-TO (LANGUAGE , ”English”) OR LIMIT-TO (LANGUAGE , ”Spanish”)) AND (LIMIT-TO (DOCTYPE , ”ar”)) AND (LIMIT-TO (AFFILCOUNTRY , ”United States”) OR LIMIT-TO (AFFILCOUNTRY , ”China”)) AND (LIMIT-TO (SUBJAREA , ”ENGI”) OR LIMIT-TO (SUBJAREA , ”MEDI”) OR LIMIT-TO (SUBJAREA , ”COMP”))

VIII. Query 2

(TITLE-ABS-KEY (data AND analysis) AND TITLE-ABS-KEY (ia)) AND PUBYEAR ¿ 2014 AND PUBYEAR ¡ 2021 AND (LIMIT-TO (LANGUAGE , ”English”) OR LIMIT-TO (LANGUAGE , ”Spanish”)) AND (LIMIT-TO (DOCTYPE , ”ar”)) AND (LIMIT-TO (AFFILCOUNTRY , ”United States”) OR LIMIT-TO (AFFILCOUNTRY , ”China”)) AND (LIMIT-TO (SUBJAREA , ”ENGI”) OR LIMIT-TO (SUBJAREA , ”MEDI”) OR LIMIT-TO (SUBJAREA , ”COMP”))