DISEÑO DE INTERFACES DE USUARIO

Actividad Bases de Datos

Santiago Montenegro Woodcock Fundación Universitaria Konrad Lorenz

Abstract—En la actualidad existen varias herramientas tecnológicas que permiten que una persona tenga acceso a una basta lista de bases de datos que contienen información relacionada con una gran cantidad de temas. Sin embargo, al realizar una búsqueda en una de estas bases de datos, esta nos puede retornar una cantidad excesiva de artículos.

En respuesta a lo anterior, se han creado métodos y herramientas para el análisis de datos con el fin de anexar filtros a la búsqueda a realizar, lo que por consiguiente limitará la cantidad de información ofrecida, volviéndose más específica y efectiva.

I. Introducción

Hoy en día las personas tienen acceso a una basta cantidad de fuentes de información gracias a las bases de datos que podemos encontrar en internet, cada una con una gran cantidad de libros, artículos, revistas; que podemos usar para indagar sobre uno o varios temas.

Sin embargo, a la hora de realizar una búsqueda, las bases de datos nos pueden ofrecer una gran cantidad de resultados de los cuales no es fácil saber si nos servirán para dar respuesta al tema de interés por parte del usuario.

En respuesta a eso, se han desarrollado diferentes tipos de análisis de datos, cada uno enfocado en diferentes objetivos para facilitar la filtración de información obtenida a la hora de realizar una investigación.

En este artículo, realizaremos un proceso de análisis de datos con el fin de obtener los mejores documentos relacionados con el tema de "Free Software" o "Software Libre" como concepto contenidos en la base de datos "Scopus" mediante el apoyo de una librería para el lenguaje de programación "Python".

II. RESULTADOS

En esta sección podremos encontrar la documentación sobre los procesos que se siguieron para realizar un análisis de datos teniendo como base el tema de "free software", iniciando con una cantidad de 61229 documentos proporcionados por la base de datos "Scopus" y concluyendo con 3 artículos [1], [2], [3], los cuales se seleccionaron de acuerdo al propósito de la búsqueda.

A. Flujo del Analisis

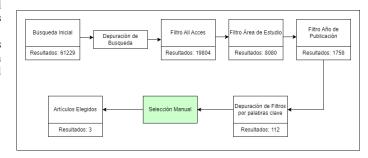


Fig. 1. Query Exploratoria.

B. Query Exploratoria

Para crear la "Query Exploratoria" tomo en cuenta las palabras principales que deberán encontrarse en los artículos a buscar (free AND software), limitando los resultados a solo los que sean del tipo "open access" para el caso de ser elegidos ciertos artículos pueda tener acceso al contenido completo. Continuado con el filtrado, limito los resultados a los que estén relacionados con las áreas de estudio sobre informática (computer science), ingeniería (engineering) y matemáticas (mathematics), y concluyendo con un filtro de año de publicación (2023, 2022 y 2021) asegurando que la información que contengan los artículos esté actualizada. Finalizado con la query representada en la Fig. 2.

$$\label{eq:tilde-decomp} \begin{split} & \text{TITLE-ABS-KEY} (\text{free AND software}) \text{ AND (LIMIT-TO (OA,"all")) AND (LIMIT-TO (SUBJAREA,"COMP") OR \\ & \text{LIMIT-TO (SUBJAREA,"ENGI") OR LIMIT-TO (SUBJAREA,"MATH"))} \\ & \text{AND (LIMIT-TO (PUBYEAR,2023) OR LIMIT-TO (PUBYEAR,2021))} \end{split}$$

Fig. 2. Query Exploratoria.

C. Analisis de Titulo

Al analizar la gráfica de la Fig.3, podemos observar que palabras de peso en la investigación como "analysis", "data" y "model" están presentes en una buena cantidad de los artículos. Sin embargo, también podemos observar palabras relacionadas al aprendizaje como lo son "study" y "learning" no se quedan atrás estando cerca de la cantidad de apariciones de las palabras anteriormente mencionadas.

En este caso, ya que la investigación se centra en "free software" de un modo conceptual, tomaremos como palabras de filtro a "análisis", "model" y "learning".

DISEÑO DE INTERFACES DE USUARIO 2

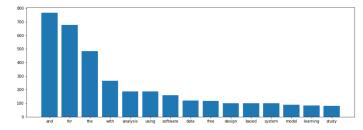


Fig. 3. Gráfica titulos.

D. Analisis de Keyword

Estudiando el resultado del análisis de las "key words" de los artículos (Fig.4), podemos observar que palabras como "system", "open source" y palabras relacionadas con tecnologías en desarrollo actuales como "artificial intelligence" y "machine learning" hacen su aparición en varios artículos.

Teniendo en cuenta lo anterior, tomaremos las palabras "open source" y "machine learning"



Fig. 4. Gráfica palabras clave.

E. Analisis de Abstract

Si analizamos la gráfica representada en la Fig.5, podremos ver que la mayoría de los artículos hacen referencia a información del año 2021 siguiéndole de cerca la información del año 2022; sin embargo se seguirán tomando en cuenta ambos años. A su vez, que dicha información en su mayoría usa el concepto de la "data" en su contenido.

Además, podemos observar palabras que repiten su aparición de los anteriores análisis como lo es "model", "analysis" y "study", por lo que se tendrán en cuenta para el query final.

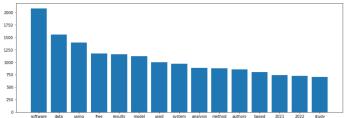


Fig. 5. Gráfica resumenes.

F. Ranking de plabras clave para la construction del nuevo filtro

Teniendo en cuenta los análisis anteriores y las palabras clave resultantes en cada uno de ellos, se van a tomar en cuenta para el query final las palabras:

- Analysis
- Model
- Learning
- · Open source
- Study
- Data

G. Query Resultante

Con el fin de complementar la query inicial, agregamos a la búsqueda las palabras clave seleccionadas anteriormente. Concluyendo con la query representada en la Fig.6

$$\label{eq:total_continuity} \begin{split} & \text{TITLE-ABS-KEY} \ (\text{ free AND software }) \ \text{AND (analysis) AND (model) AND (learning) AND (open source) AND (study) AND (data) AND (LIMIT-TO (OA,"all")) AND (LIMIT-TO (SUB)AREA,"COMP") OR LIMIT-TO (SUB)AREA,"FAGI") OR LIMIT-TO (BUB)AREA,"MATH")) AND (LIMIT-TO (PUBYEAR, 2023) OR LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PU$$

Fig. 6. Query resultante.

III. CONCLUSIONES

Para concluir, podemos afirmar que el método aplicado para el análisis de datos fue efectivo para limitar los resultados obtenidos sobre la búsqueda principal.

Además, sobre dicha búsqueda se puede decir que existe una gran cantidad de información sobre el tema de "free software" relacionado en temas de aprendizaje y para ser aplicado en las nuevas tecnologías como lo es el "machine learning".

REFERENCES

- [1] D. G. Widder, D. Nafus, L. Dabbish, and J. Herbsleb, "Limits and possibilities for "ethical ai" in open source: A study of deepfakes," in *ACM International Conference Proceeding Series*, 2022, pp. 2035–2046, cited By :1. [Online]. Available: www.scopus.com
- [2] I. Weber and C. R. Oehrn, "Nolitia: An open-source toolbox for non-linear time series analysis," Frontiers in Neuroinformatics, vol. 16, 2022, cited By :1. [Online]. Available: www.scopus.com
- [3] G. Robles, A. Capiluppi, J. M. Gonzalez-Barahona, B. Lundell, and J. Gamalielsson, "Development effort estimation in free/open source software from activity in version control systems," *Empirical Software Engineering*, vol. 27, no. 6, 2022, cited By :1. [Online]. Available: www.scopus.com