

Universidad del Quindío
Programa de Ingeniería de Sistemas y Computación
Análisis de algoritmos en el contexto de la bibliometría
Descripción inicial del proyecto

1. Introducción

La bibliometría es una disciplina que permite explorar y analizar volúmenes de datos derivados de la producción científica utilizando métodos cuantitativos y cualitativos. Se fundamenta en las matemáticas y la estadística, para establecer descripciones, relaciones, inferencias y presentaciones de la información suministrada por publicaciones científicas en diferentes áreas del conocimiento.

En bibliometría se pueden identificar diferentes indicadores. Algunos de ellos se enfocan en la productividad de los autores, índices de impacto, países, tópicos según el área de conocimiento, relación visual a partir de diferentes variables bibliométricas y colaboración entre autores.

2. Fuentes de información

La universidad del Quindío cuenta con bases de datos científicas disponibles en: <https://library.uniquindio.edu.co/databases> y las cuales agrupa por facultades. Algunas de las bases de datos son: ACM, SAGE y ScienceDirect. Cada una permite métodos de consulta, acceso y exportación de información. En este último aspecto, existen los formatos RIS, BibTex, CSV, texto plano, entre otros. Las bases de datos disponibles presentan diversas tipologías de productividad científica (artículos, conferencias, capítulos de libro, entre otros). Cada base de datos presenta limitantes en cuanto al acceso a la información y la calidad de los datos relacionados a la completitud. Para el proyecto del curso de análisis de algoritmos se plantea un dominio de conocimiento: La inteligencia artificial generativa. La cadena de búsqueda será: "**generative artificial intelligence**".

3. Propósito del proyecto

Implementar algoritmos que permitan el análisis bibliométrico y computacional sobre un dominio de conocimiento a partir de las bases de datos disponibles en la Universidad del Quindío. El desarrollo del proyecto se fundamentará en requerimientos funcionales que contemplan la implementación de diversas técnicas bibliométricas y tipos de algoritmos. Para el proyecto se es necesario el despliegue de la aplicación con la correspondiente documentación.

A continuación se hace una descripción de los requerimientos funcionales del proyecto.

Requerimiento 1. Automatización de proceso de descarga de datos.

Se debe automatizar la información de descarga sobre dos bases de datos. Posteriormente se debe unificar la información en un solo archivo garantizando una sola instancia del producto, es decir, si se identifica un producto repetido por su nombre, se debe tener un solo registro de este. El archivo unificado debe contener toda la información para cada uno de los campos (autores, título del trabajo, palabras clave, resumen, entre otros). El proceso de unificación debe ser totalmente automático tanto desde la búsqueda hasta la generación de un solo archivo

En el otro archivo se debe almacenar toda la información con el registro de los productos repetidos (artículo, conferencia, entre otros) y los cuales fueron eliminados por aparecer repetidos.

Requerimiento 2. Se deben implementar cuatro algoritmos de similitud textual clásicos (distancia de edición o vectorización estadística) y dos con modelos de IA. El análisis de cada algoritmo se debe presentar con explicación detallada paso a paso del funcionamiento matemático y algorítmico. La aplicación deberá permitir seleccionar dos o más artículos, extraer el abstract y realizar el análisis de los diferentes algoritmos de similitud textual.

Requerimiento 3. Dadas la categoría (Concepts of Generative AI in Education) y sus palabras asociadas (ver tabla abajo), se debe calcular y presentar la frecuencia de aparición teniendo como fuente el abstract de cada artículo. A continuación se debe usar un algoritmo que analice todos los abstract y genere un listado de palabras asociadas (máximo 15) de forma que se pueda mostrar la frecuencia de aparición. Finalmente debe determinar qué tan precisas son las nuevas palabras.

Categoría	Palabras asociadas
Concepts of Generative AI in Education	Generative models Prompting Machine learning Multimodality Fine-tuning Training data Algorithmic bias Explainability Transparency Ethics Privacy Personalization Human-AI interaction AI literacy Co-creation

Requerimiento 4. Implementar tres algoritmos de agrupamiento jerárquico para construir un árbol (dendrograma) que represente la similitud entre abstract científicos relacionados con el resultado de la automatización. Se debe realizar un preprocesamiento del texto (transformar el abstract), el cálculo de la similitud, la aplicación de clustering y la representación mediante un dendrograma. Es necesario determinar cuál de los algoritmos produce agrupamientos más coherentes.

Requerimiento 5. Para el análisis visual de la producción científica se debe: (1) mostrar un mapa de calor con la distribución geográfica de acuerdo con el primer autor del artículo, (2) Mostrar una nube de palabras: términos más frecuentes en abstracts y keywords. Esta nube de palabras debe ser dinámica en la medida que se adicionen más estudios al documento, (3) mostrar una línea temporal de publicaciones por año y por revista, (4) exportar los tres anteriores a formato PDF.

Requerimiento 6. El proyecto debe estar desplegado y soportado con la documentación técnica para cada uno de los requerimientos.

Documento final:

El proyecto debe estar soportado en un documento de diseño con la correspondiente arquitectura. Se debe presentar para cada requerimiento una explicación técnica con detalles de implementación. El uso de IA debe estar debidamente fundamentado y se proporcionará un documento con los aspectos que deben ser considerado.

Nota: En caso de ser necesario, la presente descripción del proyecto puede ser modificada para efectos de dar mayor claridad en su especificación. En particular con los requerimientos funcionales los cuales podrán ser explicados con mayor detalle.