**INTRODUCCION**

El algoritmo de búsqueda de secuencias maximales sirve para comparar documentos de textos dada una colección, dando como resultados un grado o porcentaje de similitud entre ellos.

En la actualidad, la minería de datos se refiere al proceso de descubrir conocimiento a partir de datos estructurados o de bajo nivel que se encuentran en una gran colección de datos, utilizando una serie de algoritmos y análisis de datos que tengan un coste computacional razonable. Así mismo, existe la minería de textos que consiste en la extracción de conocimiento a partir de datos no estructurados o lenguaje natural, usando herramientas y técnicas para el descubrimiento de patrones que nos permitirán obtener la información más importante de un texto.

Existen algoritmos de descubrimiento de patrones en colecciones de texto. Los principales están basadas en métodos a priori y en crecimiento de patrones (*pattern-growing*). Estos métodos se diferencian en que uno, el método a priori, busca candidatos a secuencias potencialmente frecuentes y luego comprueba si son frecuentes en la base de datos, y el otro, utilizan patrones que van creciendo a partir de ítems frecuentes reduciendo la búsqueda a una parte de la base de datos.

**Explicacion teorica**

Definicion del problema

Lenguaje de programación y plataforma

Parseado de los Documentos

Construccion de estructura de datos (tabla hash de palabras que apuntan a listas enlazadas de nodos de palabras)

Algoritmo de Busqueda de Secuencias Maximales

Algoritmo de almacenamiento con comparacion

**Explicación practica**

**Ventajas**

**Desventajas**