



Profundiza más

Recurso de Profundización

Síntesis de Tópicos de Textmining

Ideas principales.

1. Recuperación de Información

Encontrar **material no estructurado** que satisface una **necesidad de información** dentro de largas **colecciones de textos**
2. Sistemas de Recuperación de Información
3. Modelo Booleano

Buscador **devuelve** todos los documentos que satisfagan la una expresión **booleana**
4. Índice Invertido

Mapea las partes del documento donde el término fue encontrado, consta del **diccionario de términos** y el **posting list**
5. Análisis de Sentimientos

Recuperar sentimiento de los **textos** usando técnicas de **text mining** y **NPL**
6. Normalización

Consiste en **reducir** la aleatoriedad en una porción de texto.

Fuente: Elaboración propia

Stemming



Profundiza más



Stemming

Es el proceso de **reducir palabras a su forma raíz**. Proceso **basado en reglas**



Manag

Manag

Manag

Manager

Management

Managing



Over-sampling

universal

university

universe

El *stemming* puede ser una herramienta poderosa para reducir las palabras a su forma raíz. Sin embargo, puede incurrir en problemas de *over sampling*, tal y como se muestra en el ejemplo.

Fuente: Elaboración propia

Desarrollar un análisis de sentimientos con Python.

1. Descargar los siguientes archivos:
 - a. [tweet data](#)
 - b. [4.Modulo 4 Textmining.ipynb](#)
2. Seguir este video [video1876663457.mp4](#) desde el minuto 1:07:00 hasta el minuto 1:52:00 y hacer todas las actividades que en él se detallan