Que es un n-grama?

Es una secuencia de n ítems de una muestra de texto

Que relación hay entre Text Mining y NLP?

NLP es una herramienta utilizada para realizar text mining. El objetivo es, convertir texto en datos para análisis, via la aplicación de natural language processing (NLP) y métodos analíticos.

Que es una expresión regular y para que sirve?

Secuencia de caracteres que forman un patron de búsqueda. Utilizado para búsqueda de patrones de cadenas de caracteres u operaciones de sustitución.

Escriba una lista de procedimientos para normalización de texto típica

Pasar caracteres a minúsculas

Tokenizar

Remover stopwords

Realizar algún filtrado necesario con expresión regular

Aplicar lematización

Aplicar Steamming

Porque es necesario hacer Stemming al procesar un texto?

Para aprovechar de mejor forma los recursos al generar un solo vector de palabra y no varios dado que stemming obtiene la raíz de estas palabras.

Que es NLP?

Es un campo de las ciencias de la computación, inteligencia artificial y lingüística que estudia las interacciones entre las computadoras y el lenguaje humano. Formula e investiga mecanismos eficaces computacionalmente para la comunicación entre personas y maquinas.

Que son los StopWords en Text Mining?

Son palabras que en la gramática se repiten muchas veces y aportan muy poca información. Por ejemplo en el lenguaje español todos los artículos pueden ser stopwords. El, la, los, ellos, ellas. Etc.

Que es un corpus y que ventajas tiene sobre un vector de strings común?

Es una colección de documentos que tiene el contenido a analizar y metadata de este contenido. Contener metadata es lo que lo hace mejor que un vector simple de cadenas.

([0][1-9]|[1-2][0-9]|30)00(1[1-9][1-9][0-9]|[2-7][0-9][0-9][0-9]|8[0-9][0-6][0-9]|8970)

Que es Text Mining?

Es el proceso de obtener información útil de un texto.