

Introducción a Learning Analytics con
ejemplos prácticos

UD 04. Anexo I - Herramientas para el análisis de texto

Autor: Sergi García Barea

Actualizado Octubre 2021

Licencia




Reconocimiento – NoComercial - CompartirIgual (BY-NC-SA): No se permite un uso comercial de la obra original ni de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original.

Nomenclatura

A lo largo de este tema se utilizarán distintos símbolos para distinguir elementos importantes dentro del contenido. Estos símbolos son:

 **Importante**

 **Atención**

 **Interesante**

ÍNDICE DE CONTENIDO

1. Introducción.	3
2. Herramientas online destacadas	3
3. Ejemplo en Python proporcionado “analizador.py”	3
4. KNIME y el análisis de textos	4

UD04. ANEXO I - HERRAMIENTAS PARA EL ANÁLISIS DE TEXTO

1. INTRODUCCIÓN.

En este punto vamos a hablar de herramientas que pueden ayudarnos en el análisis de texto. Enlazaremos herramientas online destacadas. Además como extra, presentamos un programa en Python realizado para el curso (con un fin didáctico) y enlazaremos información sobre cómo realizar extracción de información en textos usando KNIME.

En algunos momentos nos referiremos al término n-grama. Este término se refiere a conjuntos de palabras seguidas, generalmente referidas al conteo. Por ejemplo, “un resumen” es un 2-grama y “herramienta que permite” un 3-grama.

2. HERRAMIENTAS ONLINE DESTACADAS

En este punto describimos herramientas online destacadas para el análisis de textos:

- <https://countwordsfree.com/>
 - Herramienta que permite contar palabras, frases, líneas. Además permite ver frecuencias tanto de palabras como de n-gramas hasta $n=4$. También permite en el análisis obviar “stop words” (palabras vacías) en distintos idiomas.
 - Entre otras funciones indica el tiempo estimado de un texto y el tiempo estimado de locución del mismo, así como la posibilidad de que la web haga de locutor del mismo.
- <http://guidetodatamining.com/ngramAnalyzer/>
 - Analizador de n-gramas de un texto.
- <https://www.browserling.com/tools/remove-punctuation>
 - Herramienta para eliminar símbolos de puntuación online.
- <https://legible.es/>
 - Herramienta para obtener métricas de legibilidad de un texto.
- <https://niram.org/read/>
 - Indica el tiempo estimado de lectura de un texto.
- <https://tobloef.com/text2mindmap/>
 - Herramienta para crear mapas mentales a partir de texto.

3. EJEMPLO EN PYTHON PROPORCIONADO “ANALIZADOR.PY”

Con un fin puramente instructivo y por relacionar la temática con la programación se ha desarrollado para el curso un pequeño programa en Python 3 para realizar análisis de texto.

Este programa nos proporcionará información sobre el texto, palabras, etc. e incluso intentará realizar un resumen basado en intentar localizar de forma rudimentaria frases con mayor carga de contenido y presentarlas ordenadas.

El programa está disponible en “**AnalizadorTextos.zip**”. Incluye comentarios y requiere la instalación de la biblioteca PDFMiner con “pip install pdfminer” y “pip install pdfminer.six”.

Puede utilizarse para:

- Extraer información de ficheros txt. Se usa ***“python analizador.py fichero.txt”***
- Extraer información de ficheros PDF. Se usa ***“python analiazar.py fichero.pdf pdf”***
 - El último parámetro (“pdf”) es para indicar que es ese tipo de fichero.

4. KNIME Y EL ANÁLISIS DE TEXTOS

KNIME es ampliamente utilizado para el análisis y extracción de información de textos.

Para el uso de las mismas, deben instalarse las siguientes extensiones de KNIME:

- “KNIME textprocessing”
- “KNIME textprocessing Spanish Language Pack”

En esta unidad no hemos tratado el uso de KNIME como herramienta de análisis de textos, pero si quieres saber más al respecto, tienes unos tutoriales aquí:

- <https://www.youtube.com/watch?v=E0p5p90onDA>
- <https://files.knime.com/sites/default/files/Text%20Mining%20Course%20for%20KNIME%20Analytics%20Platform.pdf>