

Análisis de algoritmos

Taller 2

Angie Tatiana Peña Peña

Juan David Aldana Rincon

Andrés David Rueda Sandoval

Carlos Andrés Borja González

Santiago Guerrero Danderino

Pablo Arturo Rodríguez González

2023

**Introducción:**

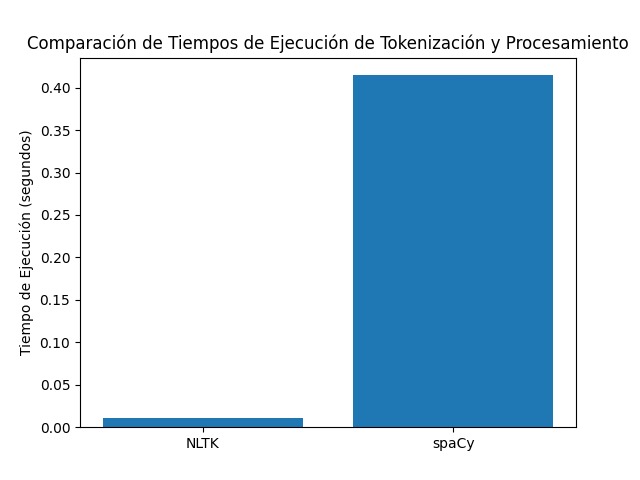
El siguiente taller consiste en el análisis de los siguientes videos de Youtube:

* <https://www.youtube.com/watch?v=fxXU1vvbQyA>
* <https://www.youtube.com/watch?v=6_Cs1tKzzbE>
* <https://www.youtube.com/watch?v=uLC0dJhQ1d0&t=3775s>

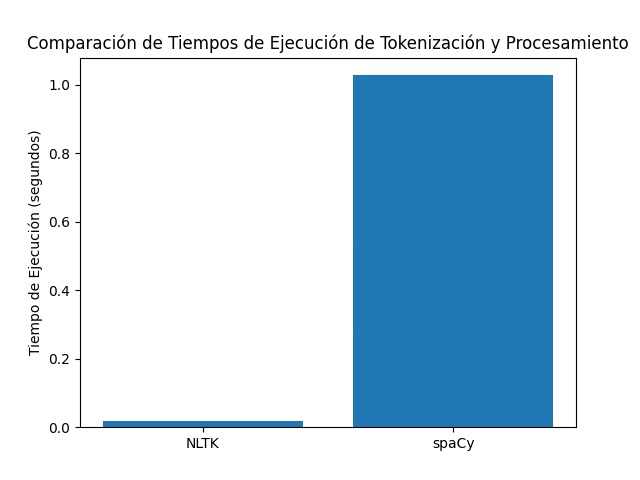
Gracias a un script en Python en el cual permite realizar una extracción de datos y de palabras clave de cada uno de los videos sobre debates entre los candidatos a la alcaldía de Bogotá, y realizando una limpieza de datos de la extracción, para lograr eliminar conectores y palabras sin sentido dentro de los resultados, se pudieron obtener las siguientes gráficas y nubes de palabras.

**Gráficos comparativos:**

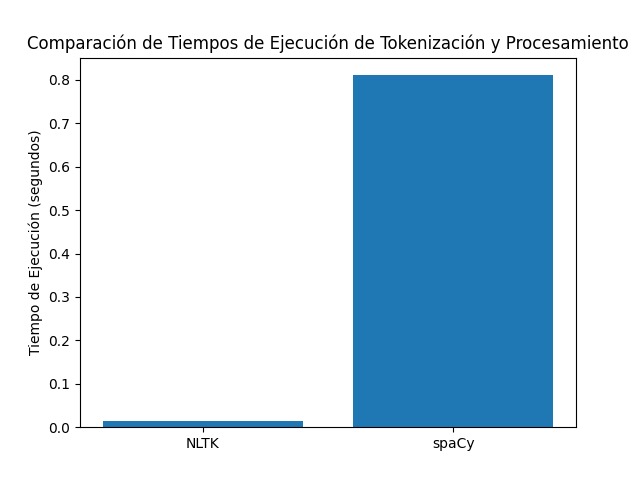
* <https://www.youtube.com/watch?v=fxXU1vvbQyA>



* <https://www.youtube.com/watch?v=6_Cs1tKzzbE>

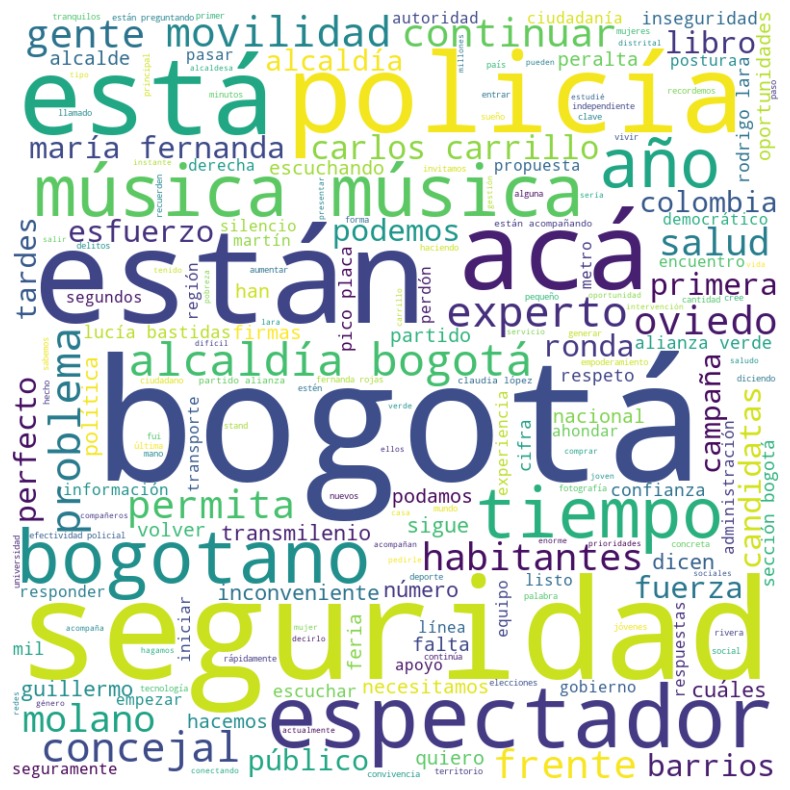


* <https://www.youtube.com/watch?v=uLC0dJhQ1d0&t=3775s>



**Nubes de palabras:**

* <https://www.youtube.com/watch?v=fxXU1vvbQyA>



* <https://www.youtube.com/watch?v=6_Cs1tKzzbE>



* <https://www.youtube.com/watch?v=uLC0dJhQ1d0&t=3775s>



**Reflexiones y observaciones:**

* En cuanto a los gráficos comparativos, podemos concluir que la biblioteca NLTK en cuanto a tiempos de ejecución, es mucho más eficiente que la biblioteca spaCy para el análisis y procesamiento de textos debido a que spaCy tiene tiempos de ejecución muy altos comparados a NLTK.
* En cuanto a las nubes de palabras generadas, podemos concluir que los temas que más se tocan en estos videos de debates, obviando a “Bogotá”, son: problema, mujeres, obras, educación, metro, policía, seguridad y Transmilenio.
* Por lo anterior se puede concluir que, la abstracción de datos y la limpieza de los mismos, sirvieron para obtener resultados más certeros y con temas más relacionados al debate.