

This project is about finding character names among the top ten most common meaningful words in the text. First, I gathered data from a provided URL using a Python web-scraping tool called BeautifulSoup. Then I have used the tokenization technique with a regex tokenizer to tokenize words only from the text. Moreover, I have converted the words into lowercases using the `.lower()` method, and finally filtered stopwords using the corpus of the natural language toolkit (NLTK). The main result of the project is that I found that there are 4 names out of the ten most common words of the text, are protagonists' names.