The main purpose of this project is to build a recommendation system for Charles Darwin’s books, which will determine which books are closed to each other based on the similarity of the discussed topics. In this project, multiple NLP techniques were used. First, I used the Tokenization method to tokenize the text. As Darwin was using different words to refer to a similar concept, a stemming process should be used to return these words to their roots and also not to have many words that have the same meaning. Also, a bag-of-words(BoW) model was created to represent the books as a list of all unique tokens they contain associated with the respective number of occurrences. After converting the results from the BoW model into a dataframe, the most common words were “have”, “on”, “it”, ...etc. These are just stopwords and it is not useful for our investigation. The next step was about using a TF-IDF model to get rid of the stopwords. The model worked almost perfectly and gave us the most frequent words in the text. Lastly, a cosine similarity was used to measure how books are related to each other. As a final result of the project, we were able to know the books that are similar to the “On the Origin of Species” book. The results can be mainly used for book stores to make recommendations for Charles Darwin’s readers.