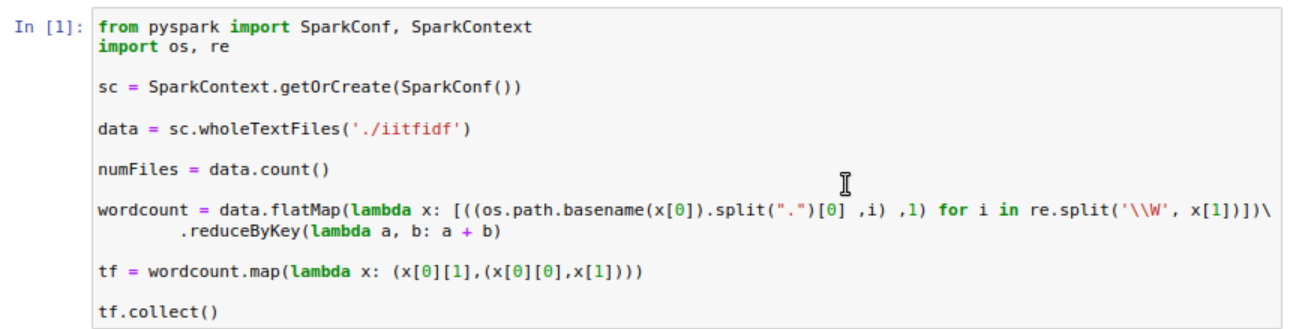
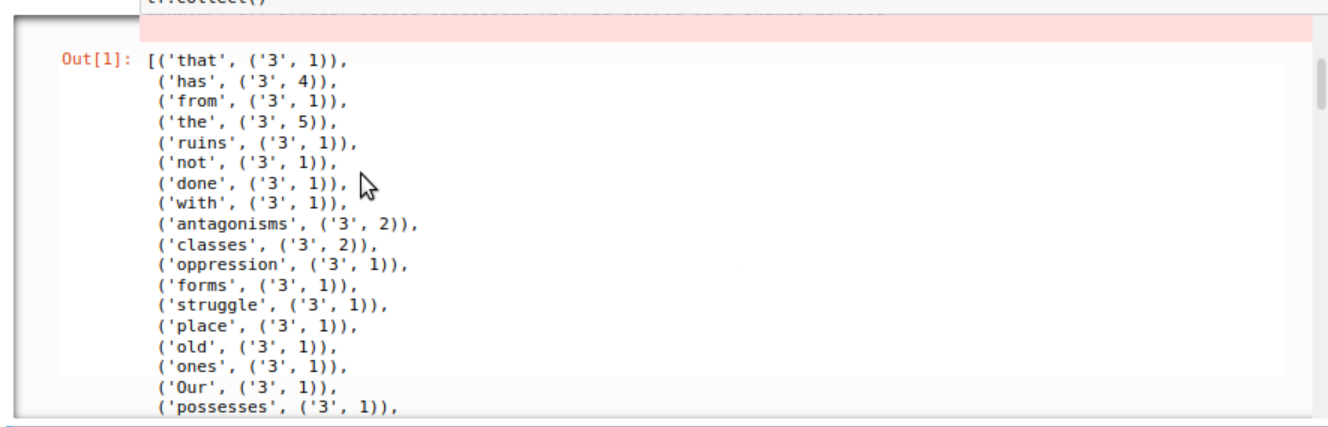
**Basic**

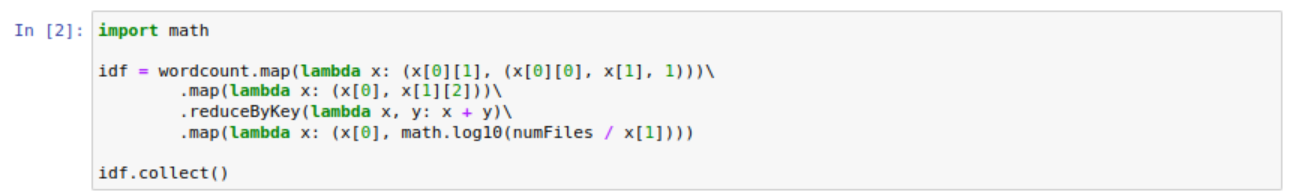




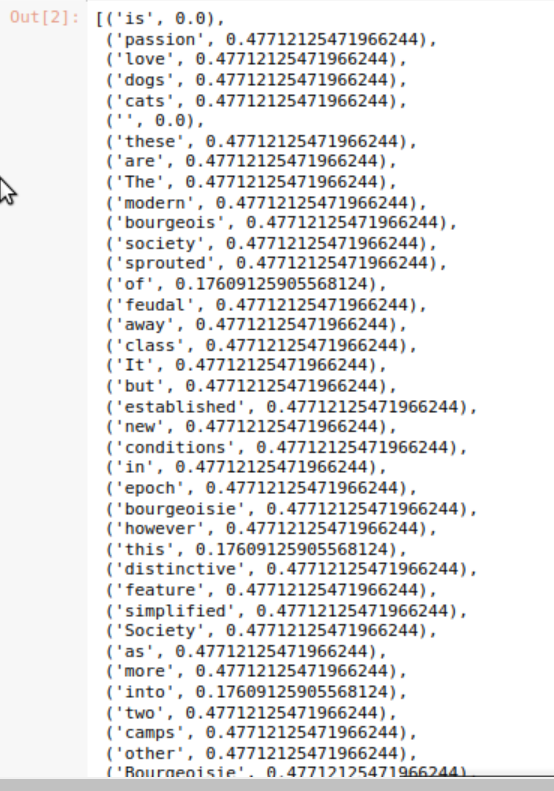
**Calculate the term frequency of the each word in documents**



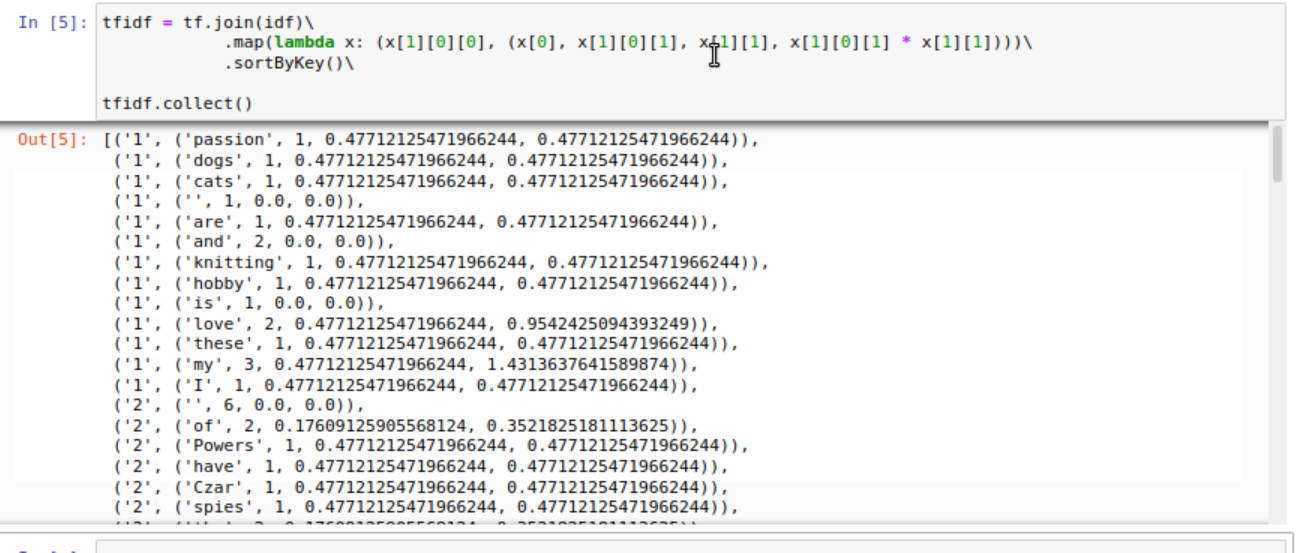
**The term frequency output**



**Calculate the inverted document frequency**

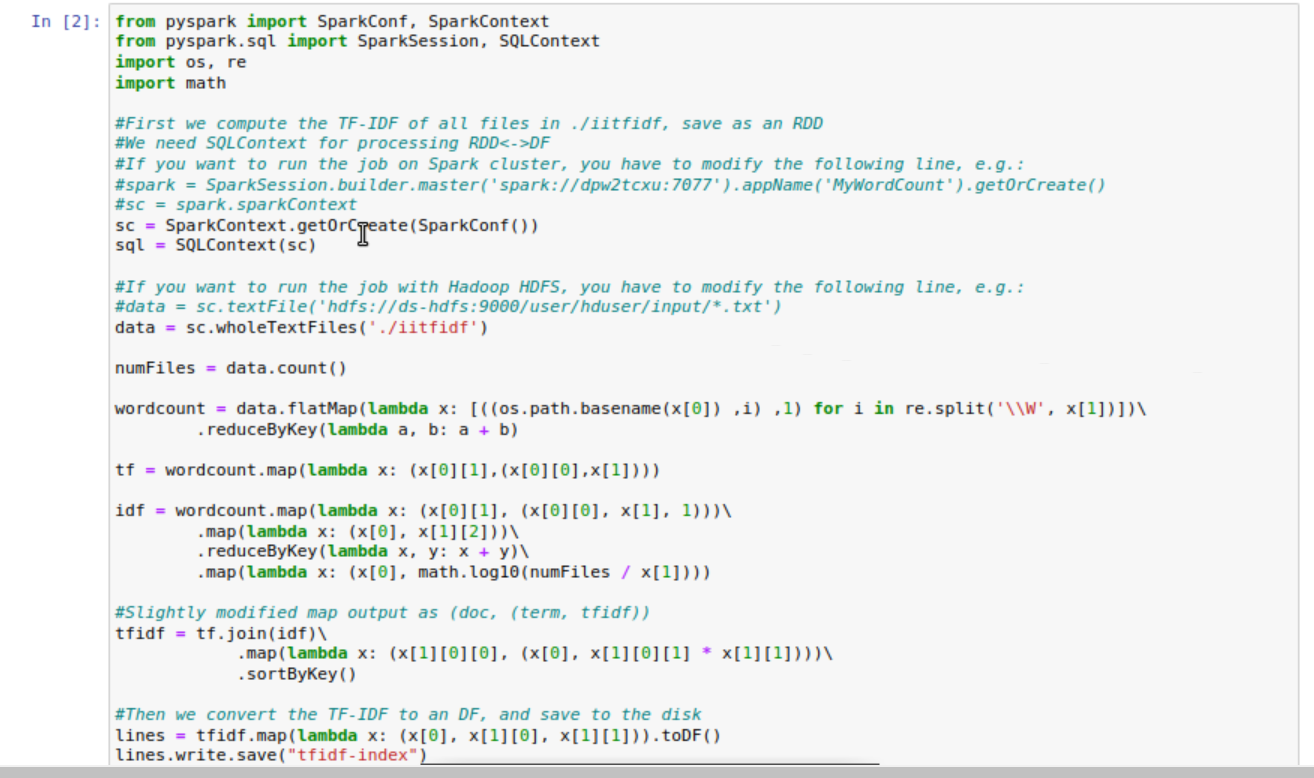


**The output of the inverted document frequency**



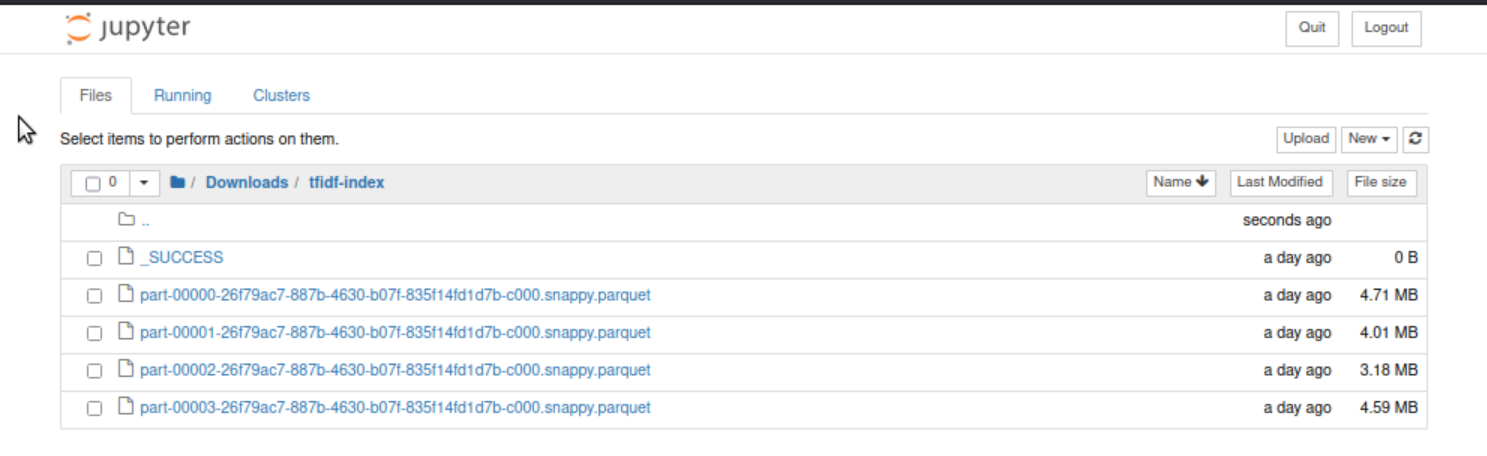
**Computing TF-IDF and show the result**



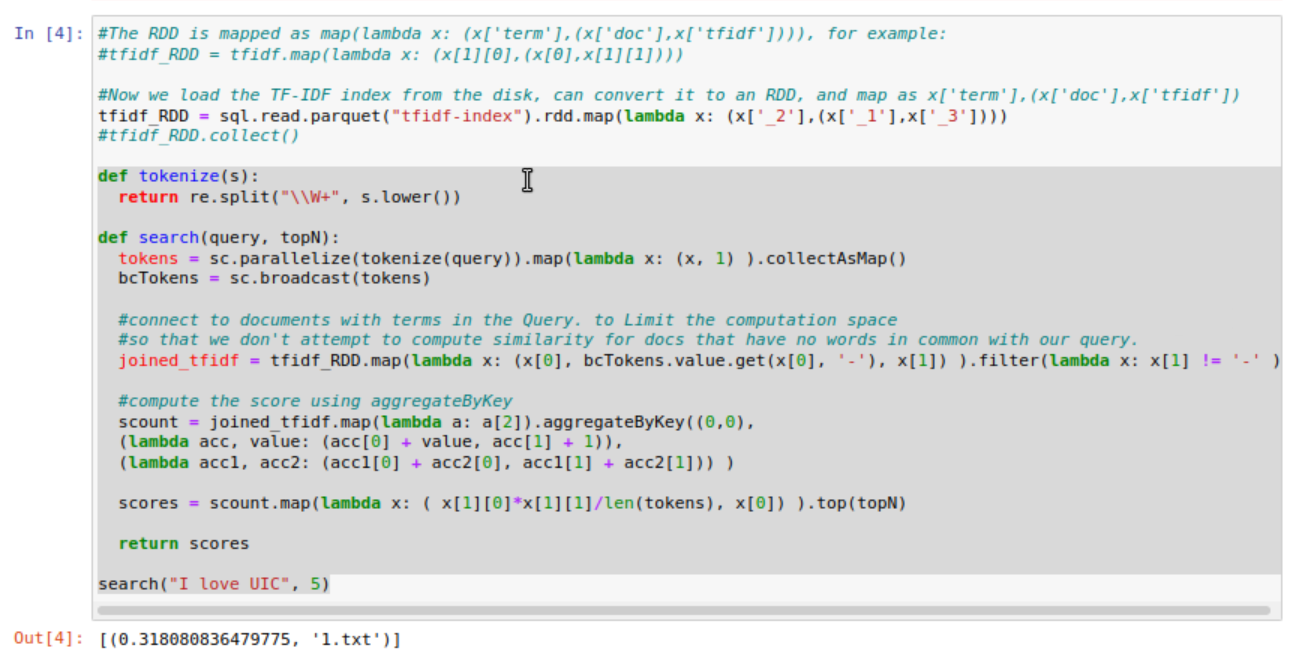


**Based on the given English documents, calculate the tf-idf and save in a directory.**



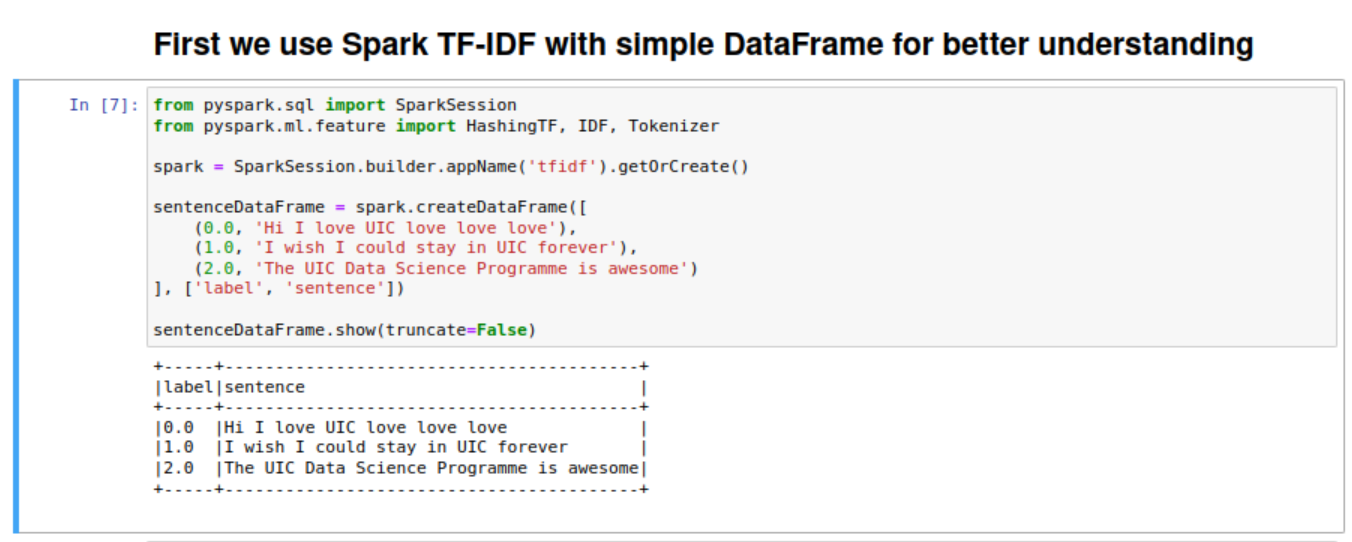


**Here is the files that generated by the codes**

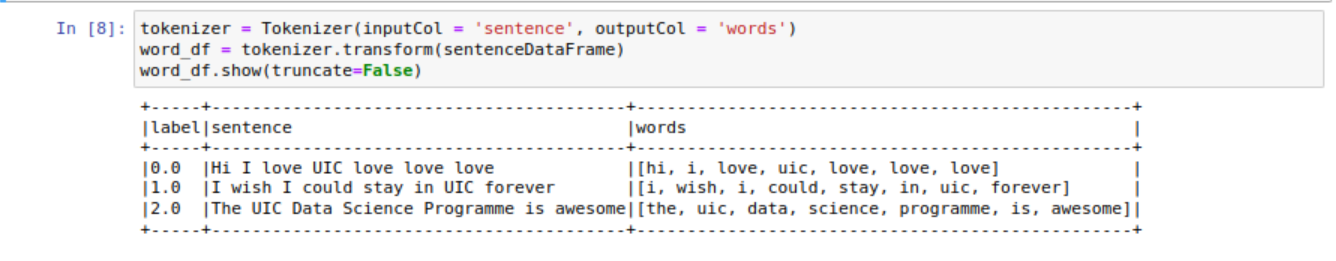


**Search specific key words based on the generated tf-idf and return the most similar document and its score.**

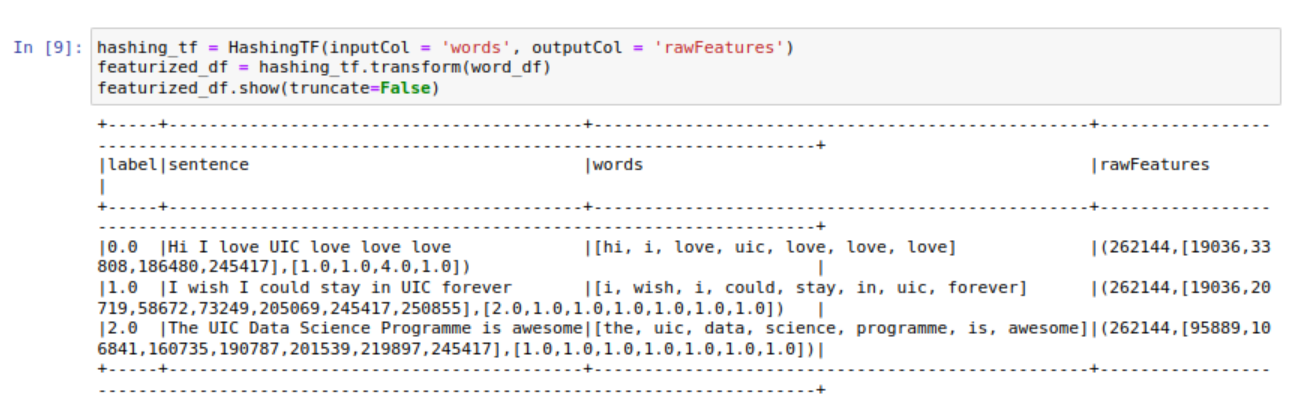




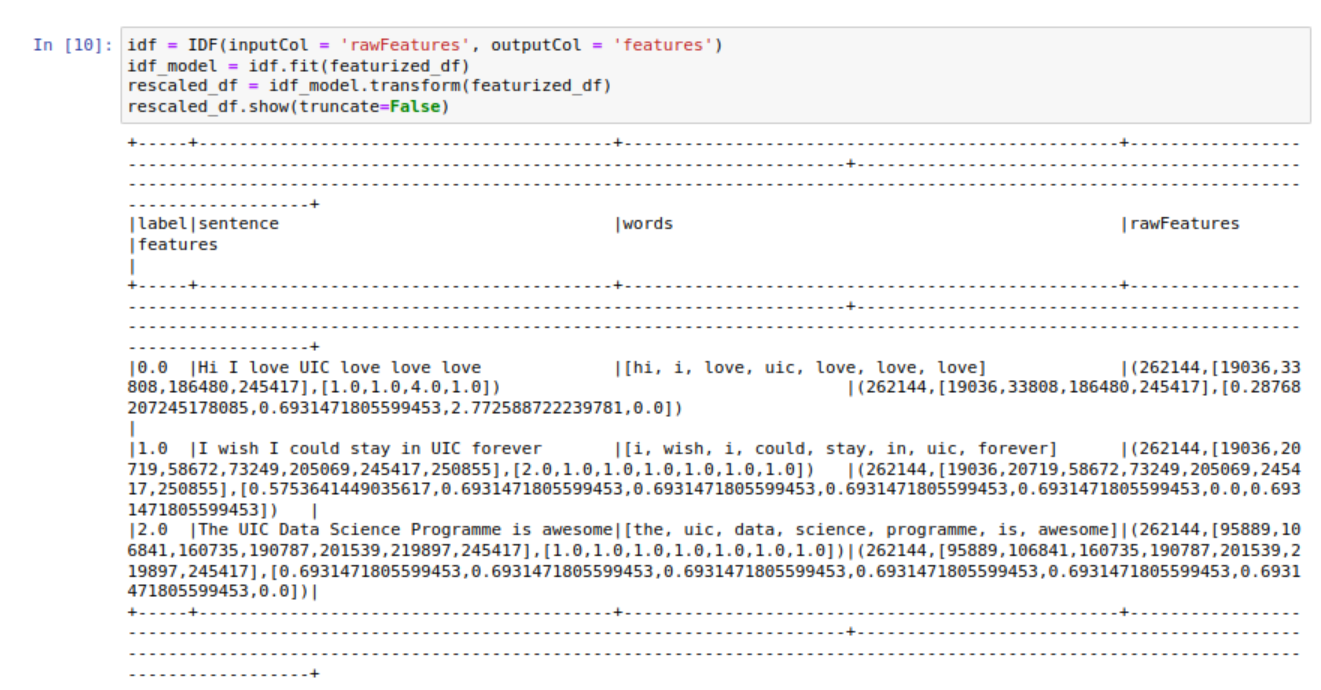
**Use Spark TF-IDF with simple data frame**



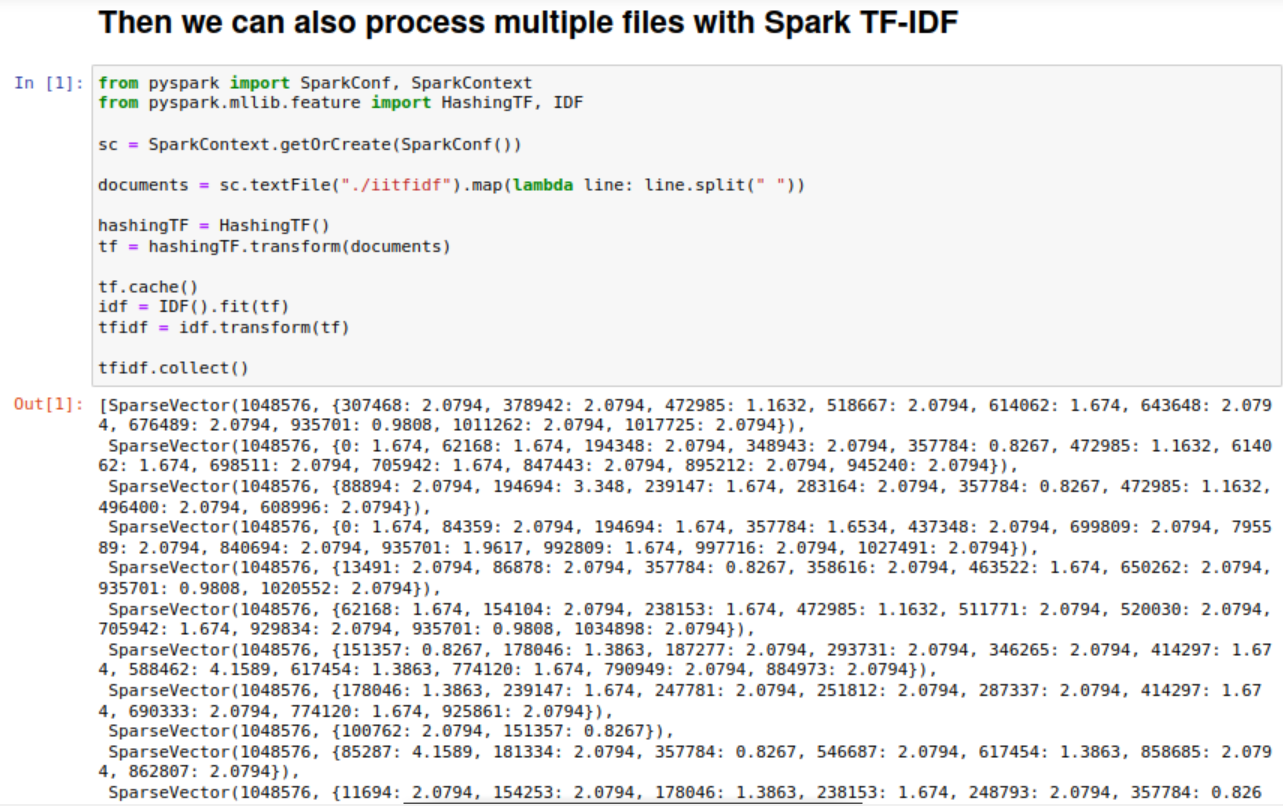
**Tokenize the word**



**Encode the words and calculate term frequency**



**Calculate the inverted term frequency**



**Process multiple files with Spark TF-IDF**