IR Assignment 1

Aakash Khadka, Gowtham Premkumar, Max Neubauer

1.a)
In order to complete the assignments in a group we used GitHub for version control. The project is a Maven project using Java 8 and Lucene 7.4.
We included the following Lucene packages:

- lucene-core
- lucene-highlighter
- lucene-queryparser
- lucene-analyzers-common

For the first task we used a function that accepts a Tokenizer Object and a String. The String is put into the Tokenizer. From the Tokenizer we define an Attribute that holds all the Tokens. Then we iterate over the Tokens and print them.

```
C:\Users\laptopuser\.jdks\corretto-1.8.0_432\bin\java.exe ...
Standard Tokenizer:
Today,is,sunny,She,is,a,sunny,girl,To,be,or,not,to,be,She,is,in,Berlin,today,Sunny,Berlin,Berlin,is,always,exciting,
Whitespace Tokenizer:
Today,is,sunny.,She,is,a,sunny,girl.,To,be,or,not,to,be.,She,is,in,Berlin,today.,Sunny,Berlin!,Berlin,is,always,exciting!,

Process finished with exit code 0
```

The Whitespace Tokenizer only trims white spaces. Standard Tokenizer also gets rid of punctuation.

1.b)
The stop words are put into a CharArraySet. Then we use a TokenStream with the Standard Tokenizer and the stop words. At last we use the same mechanic as above to iterate over the tokens.

```
Standard Tokenizer:
Today,is,sunny,She,is,a,sunny,girl,To,be,or,not,to,be,She,is,in,Berlin,today,Sunny,Berlin,Berlin,is,always,exciting,
Analyzer and StopwordFilter:
Today,sunny,She,a,sunny,girl,or,not,She,Berlin,today,Sunny,Berlin,Berlin,always,exciting,
Process finished with exit code 0
```

1.c)
We created a CustomAnalyzer called OurAnalyzer which extends the Analyzer class. Therefore, we need to override the createComponent method. There we applied the filters. We created a StandardTokenizer, a LowercaseFilter, a StopwordFilter and a PorterStemmerFilter.

```
Analyzer and StopwordFilter:
Today,sunny,She,a,sunny,girl,or,not,She,Berlin,today,Sunny,Berlin,Berlin,always,exciting,
Custom Analyzer
todai,sunni,she,a,sunni,girl,or,not,she,berlin,todai,sunni,berlin,berlin,alwai,excit,
Process finished with exit code 0
```