IR Assignment 2

Aakash Khadka, Gowtham Premkumar, Max Neubauer

To organize our folder structure the source code of this assignment is put into a new package.

2.a)
In order to implement Assignment 4.1 of the exercise sheet we add all documents into an IndexWriter. We have a StandardAnalyzer with lowercasefilter. Then we define the Query "sunny AND exciting". Then we define a QueryParser with our custom Analyzer. After that we use an IndexSearcher to get the top 100 matching documents. Finally we print the matching documents. In order to show that the code works accordingly we used a second query for validation.

```
There are no occurrence of the words: sunny AND exciting
Both of the terms "to AND be" occur in document: 2
```

2.b)
In order to print the postingslist we create an IndexReader object. Then we built a static method to print the postings list. In this method we loop through all the terms in the index object and check whether it matches with the target term. Then we print the target term, its total frequency and its document frequency. After that we iterate over all documents it occurs in and mention its frequency in that document and aggregate all positions in a list to display them as for the format.

```
Printing Posting List for "sunny" :
[sunny : 3 : 3] -> [ 0 : 1 : [2] ] -> [ 1 : 1 : [3] ] -> [ 4 : 1 : [0] ]
Printing Posting List for "to" :
[to : 2 : 1] -> [ 2 : 2 : [0, 4] ]
```