

```

1 import java.io.IOException;
2 import java.nio.charset.Charset;
3 import java.nio.file.Files;
4 import java.nio.file.Paths;
5 import java.util.Collections;
6 import java.util.HashMap;
7 import java.util.Map;
8
9 public class Document {
10     private Map<String, Integer> wordCount = new HashMap<String, Integer>();
11
12     public Document(String path) throws IOException {
13         countWordsFromFile(path, Charset.forName("Cp1252"));
14     }
15
16     public boolean doesTermAccur(String term){
17         return wordCount.containsKey(term);
18     }
19
20     /**
21      * calculate term frequency for a given term in document
22      * @param term: term w to calculate tf
23      * @return value for tf
24      */
25     public double get_tf(String term){
26         int f_ji = 0;
27         if (wordCount.containsKey(term)) {
28             f_ji = wordCount.get(term);
29         }
30         int max_k_fkj = Collections.max(wordCount.values());
31
32         return calculate_TF((double)f_ji, (double) max_k_fkj);
33     }
34
35     /**
36      * calculates tf
37      * @param f_ji: number of occurrences of a term
38      * @param max_k_fkj: maximum number of occurrences of any term
39      * @return value for tf
40      */
41     private static double calculate_TF(double f_ji, double max_k_fkj){
42         return f_ji/max_k_fkj;
43     }
44
45     /**
46      * Method to count words from a text file
47      * @param path: path of file to read
48      * @param encoding: encoding of file
49      * @throws IOException if file can't be loaded
50      */
51     private void countWordsFromFile(String path, Charset encoding)
52         throws IOException {
53         String text = new String(Files.readAllBytes(Paths.get(path)), encoding);
54         //Trim file in order to get only raw words
55         text = text.replaceAll("[\\s]", " ");
56         text = text.replaceAll("[^A-ZaäöüÄÖÜß ]", "");
57         text = text.toLowerCase();
58         String[] parts = text.split(" ");
59
60         //count words
61         for (String part : parts) {
62             if (wordCount.containsKey(part)) {
63                 wordCount.put(part, wordCount.get(part) + 1);
64             } else {
65                 wordCount.put(part, 1);
66             }
67         }
68     }
69 }

```