

**Gebze Technical University
Computer Engineering**

CSE 222 - 2018 Spring

HOMEWORK 06 REPORT

**Rumeysa Karakavak
141044063**

Course Assistant: Ayse Serbetci Turan

1 INTRODUCTION

1.1 Problem Definition

Implement two different HashMap classes to perform basic Natural Language Processing operations. Specifically, you will read a text dataset consisting of multiple input files and keep the words in the Word HashMap. The key for the word hashmap will be the words and the value will refer to another hashmap (file hashmap) which keeps the occurrences of the word in different files. The key for the file hashmap is the filename and the value is an arraylist containing the word positions in that file. In order to be able to traverse the word map in an efficient way, each entry should keep a pointer to the next inserted entry, allowing the structure to be Iterable (implements Iterable interface).

By this way, the methods such as containsValue(), containsKey() and keySet() can be implemented in a more efficient way, without the need for visiting empty cells in the table.

1.2 System Requirements

This program was coded and can running on IntelliJIDEA(it can runs another platforms probably). Used latest version of JDK 11.0.2. .

Operating System (i.e. Windows 10, Linux or Mac OS X 10.3.8)

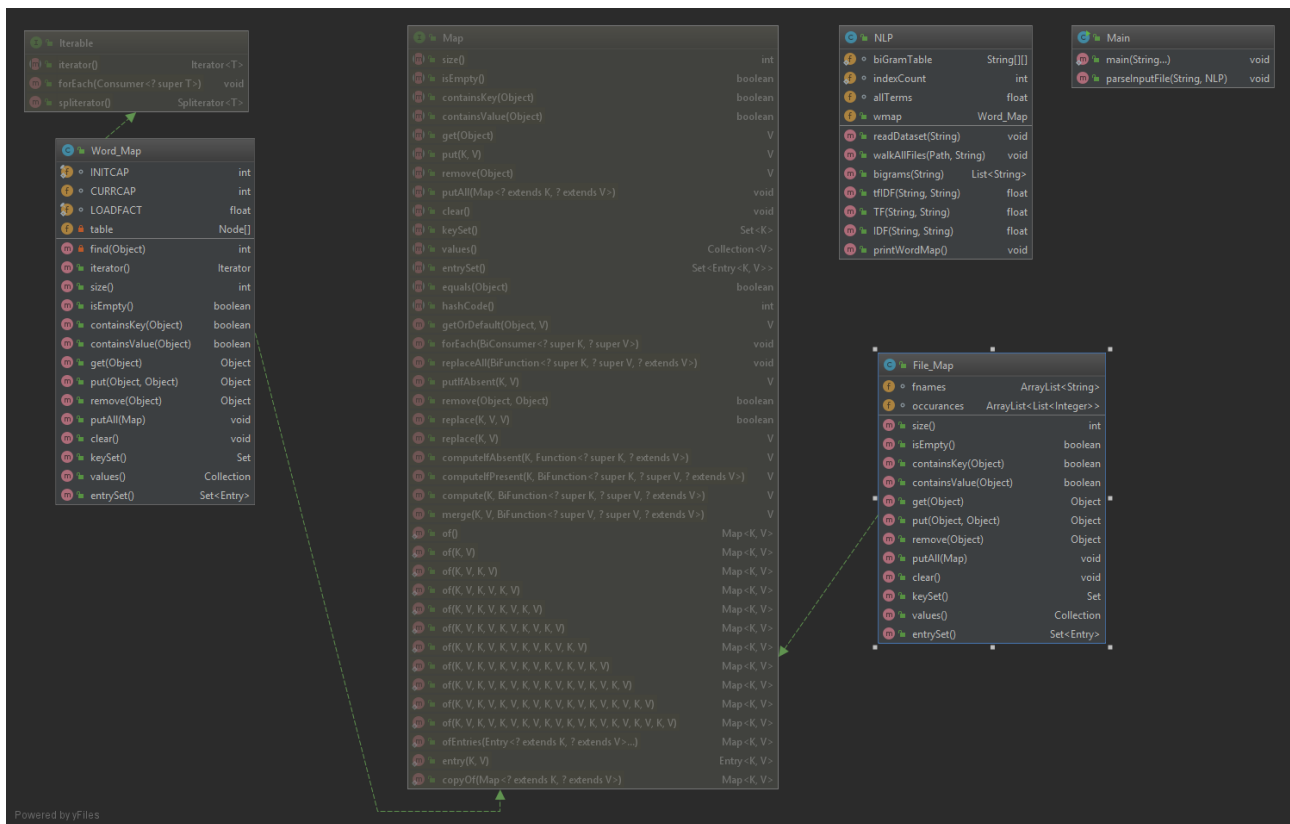
Processor Speed (i.e. Pentium 4, 3.2 GHz or Power PC G5, 2.0 GHz)

Memory, a.k.a. RAM (i.e. 512 MB)

Hard Disk Space (i.e. 80 GB available)

2 METHOD

2.1 Class Diagrams



2.2 Use Case Diagrams

How is this software supposed to be used? What is expected of the user? What is that button for? Is the user supposed to press it, click it, punch it? Explain step by step how the user(s) are supposed to/expected to use your software - with diagrams if necessary.

2.3 Problem Solution Approach

Explain your method and justify your decisions. How did you achieve the homework goal? Why did you make the design decisions that you've made? Which data structure/algorithm you are using and why? What is your program's time and space complexity? If you didn't solve the problem where did you get stuck? Why?

3 RESULT

3.1 Test Cases

Test queries:

bigram very
tfidf coffee 0001978
bigram world
bigram costs
bigram is
tfidf Brazil 0000178

3.2 Running Results

```
"C:\Program Files\Java\jdk-11.0.2\bin\java.exe" "-javaagent:C:\Program Files\JetBrains\IntelliJ IDEA 2018.3.5\lib\idea_rt.jar=63917:C:\Program Files\JetBrains\IntelliJ IDEA 2018.3"
[very difficult, very rapid, very vulnerable, very aggressive, very promising, very soon, very attractive]
0.011813279
[world markets, world bank, world as, world grain, world cocoa, world for, world market, world made, world price, world coffee, world prices, world share, world tin]
[costs and, costs Transport, costs of, costs have]
[is searching, is helping, is flowering, is willing, is scheduled, is unlikely, is beginning, is put, is sceptical, is no, is trying, is after, is passed, is some, is insufficient]
0.0019984015
Process finished with exit code 0
```