LW 2021 - Internship

1. **Java assignment**
    a. Given an input plain text file count the number of appearances for each word (assume words are separated by space and ignore case).
    List the number of appearances for each word in the form
    word1=12
    word2=112
    word3=45
    ...

    b. Sort the previous list by the number of appearances of each word in descending order and then print it.
    word2=112
    word3=45
    word1=12
    ...

    e.g For the following input: This is a random sentence for a random test using this random words.
    You get this output
    random=3
    a=2
    this=2
    sentence=1
    using=1
    test=1
    words.=1
    for=1
    is=1

    c. For the sentences in the file https://goo.gl/pTSVcQ compute the Jaccard distance (https://bit.ly/2FfoyxE) between the sentences in the file and print them.

    d. **Bonus**

        i. Use Java 8 features
        ii. What happens when the input file is 1MB/10MB/1GB/1TB?
        iii. Use the following file for test https://goo.gl/pTSVcQ . How many appearances does the word "his" have?
        iv. Use the following file for test http://norvig.com/big.txt . How many appearances does the word "was" have?
        v. Notes: We care about the way you model the algorithm, oop, iterations, data structures, optimizations.