

READ ME FILE

ΑΡΗΣ ΖΑΦΕΙΡΑΤΗΣ ΑΜ 1115201100169

Γλώσσα υλοποίησης C++

Εχει δημιουργηθεί **makefile**

Εντολή μεταγλώττισης **make**

Εντολή εκτέλεσης **./main -i smallDataset.txt -k 7**

Η 1 εργασία έχει δημιουργηθεί με **netbeans** IDE 8.2. Εχει δοκιμαστεί στα μηχανήματα της σχολής.

Γινετε σωστη διαχείριση της μνημης , δεσμευση/αποδεσμευση. Για το σκοπό αυτο δοκιμαστηκε και με **valgrind!!!**

Εχω σπασει το προγραμμα μου σε αρκετα αρχεια , προσεγγμενος και ευαναγνωστος κωδικας

Εχω δημιουργήσει 5 κλάσεις :

1. **Map:** Κύριες αρμοδιότητες το διαβάσμα του αρχείου , ελεγχος λαθων , αποθηκευση των κειμενων στον πινακα `docArray` , υπολογισμός του avgDL
2. **Trie:** Η δομή έχει υλοποιηθει **οπως ακριβως προτινεται** απο την εκφωνηση, όπως το παραδειγμα της εκφωνησης. Κυριες αρμοδιότητες ειναι η εισαγωγή λέξης στο trie, αναζήτηση , οι λειτουργίες `tf`, `df`, καταστροφή της δομης. Η καταστροφή της δομής και της λειτουργιας `df` γίνεται χρησιμοποιώντας αναδρομικες συναρτήσεις.
3. **PostingList:** Εισαγωγή δεδομένων στην Posting List , Υπολογισμός IDF, επιστροφή δεδομένων απο κάποιο συγκεκριμένο κομβο της λίστας.
4. **WordList** και 5. **ScoreList** Ειναι **σημαντικές στο κομματι της αναζητησης και υπολογισμού του σκορ**. Πιο συγκεκριμενα στην WordList αποθηκευται ενας δείκτης που δειχνει στην posting list καθε όρου (λεξης) που πληκτρολογηθηκε στην αναζητηση. Αλληλεπιδρωντας με την ScoreList η οποια αποθηκευει/ανανεωνει στην λιστα της τα σκορ καθε κειμενου που βρεθηκε η λεξη. Επισης η ScoreList έχει σαν αρμοδιότητα την εκτύπωση των δεδομενων της, και την ταξινόμηση των δεδομενων που γινεται με τον αλγοριθμο της **mergeSort** που έχει πολυπλοκότητα **$n \log n$** .

Πηγες:

<https://stackoverflow.com/questions/3501338/c-read-file-line-by-line>

<https://stackoverflow.com/questions/11391870/sscanf-writing-to-a-character-array>
<http://www.cplusplus.com/reference/cstring/strlen/>
<https://stackoverflow.com/questions/632846/clearing-a-char-array-c>
<http://www.cplusplus.com/reference/cstring/strtok/>
<http://www.cplusplus.com/forum/beginner/8388/>
<https://gist.github.com/hnguyen091188/efa43431593f2370007e>