



# UNIVERSITY OF PIRAEUS

Ονοματεπώνυμο: Σταμάτιος Ορφανός

Αριθμός Μητρώου: Ε17113

Μάθημα: Ανάκτηση Πληροφοριών

## Περιεχόμενα

1. Αρχεία .....	Σελ. 1
2. Εφαρμογή .....	Σελ. 2
3. Queries .....	Σελ. 3

## Αρχεία

Τα πιο σημαντικά σημεία για την εκτέλεση της εφαρμογής είναι τα αρχεία όπως το Ανεστραμμένο Ευρετήριο και το Mapper αρχείο. Ένα από τα βασικά πλεονεκτήματα της εφαρμογής είναι ότι χρειάζεται να οριστούν μόνο την πρώτη φορά και σε κάθε επόμενη χρήση θα φορτώνονται απλά από την μνήμη.

Φυσικά πρέπει να οριστούν κατάλληλα τα path των αρχείων αυτών για την σωστή λειτουργία της εφαρμογής. Για ευκολία χρήσης της εφαρμογής τα path των εγγράφων βρίσκονται στην βασική εφαρμογή.

```
InvertedIndex invertedIndex = new InvertedIndex();
String indexPath = "/home/stamatiosorfanos/git/SearchEngine/src/invertedIndex.ser";
File invertedIndexFile = new File(indexPath);

DocIDMapper mapper = new DocIDMapper(8001);
String mapperPath = "/home/stamatiosorfanos/git/SearchEngine/src/mapper.ser";
File mapperFile = new File(mapperPath);

if (!invertedIndexFile.exists() || !mapperFile.exists()) {
    File data = new File("/home/stamatiosorfanos/git/SearchEngine/Reuters8Kdata");
    Parser.parseData(data, mapper, invertedIndex);
    invertedIndex.write(indexPath);
    mapper.write(mapperPath);
} else {
    invertedIndex.load(indexPath);
    mapper.load(mapperPath);
}
```

## Εφαρμογή

Η εφαρμογή ξεκινάει ζητώντας από τον χρήστη να εισάγει το query, το οποίο θα περιλαμβάνει τα terms που θέλει να αναζητήσει.

```
(base) stamatiisorfanos@pop-os:~/git/SearchEngine$ cd /home/stamatiisorfanos/git/SearchEngine ; /usr/bin/env /usr/lib/jvm/java-14-openjdk-amd64/bin/java -Xmx16G -XX:+ShowCodeDetailsInExceptionMessages -Dfile.encoding=UTF-8 -cp /home/stamatiisorfanos/.config/Cod
e/User/workspaceStorage/1f103e0fbc3af7f96b6b6aa0a13ab204/redhat.java/jdt_ws/SearchEngine_6388bb4e/bin Application

Welcome to the browser
Please enter the query:
computer systems sale
1./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-000x9.txt
2./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-019x678.txt
3./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/earn/reut2-010x547.txt
4./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-010x472.txt
5./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-003x286.txt
6./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-012x320.txt
7./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-019x905.txt
8./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-001x739.txt
```

Αφού ο χρήστης εισάγει το query, τότε θα επιστραφούν τα κατάλληλα αποτελέσματα. Αμέσως μετά ο χρήστης καλείται να επιλέξει ένα από τα αποτελέσματα εισάγοντας τον αντίστοιχο αριθμό του αποτελέσματος. Σε περίπτωση όπου ο χρήστης αποφασίσει ότι δεν θέλει να δει τα αποτελέσματα, τότε μπορεί να εισάγει τον χαρακτήρα c για να συνεχίσει (continue).

```
computer systems sale
1./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-000x9.txt
2./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-019x678.txt
3./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/earn/reut2-010x547.txt
4./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-010x472.txt
5./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-003x286.txt
6./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-012x320.txt
7./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-019x905.txt
8./home/stamatiisorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-001x739.txt

Please type the number of the page you would like to visit or 'c' to continue
3
Honeywell Bull Inc ,owned by Honeywell Inc lt HON ,lt Cie des Machines Bull and lt NEC Corp ,said it expects its 1987 revenues to i
ncrease 15 to 20 pct over its current level of about 1.9 billion dlrs .Honeywell Bull president Jerome Meyer also told the press co
nference the company was profitable ,adding the company aimed to improve profits over time .Honeywell Inc earlier today received 52
7 mln dlrs in cash for the sale of 57.5 pct of its computer business to Bull and NEC .Honeywell will retain a 42.5 pct interest in
Honeywell Bull and Bull will own an equal amount .NEC owns 15 pct of the company .Meyer ,who had been executive vice president of H
oneywell Information Systems ,said Honeywell Bull would work closely with NEC and Bull in both developing new products and marketin
g computers to multi-national companies ."We have n't been growing as fast as we 'd like ,but we are going to turn that around ,"Me
yer said .He said business was soft in the U.S. He said the company was reorganizing its distribution programs and its staff ,reass
igning and laying off selected employees .He also said the company was hiring new workers .Honeywell Bull chairman Jacques Stern ,w
ho is also chairman of Bull ,said he believed the computer industry could be a fast growing market for companies that provide the t
ype of computers customers want ."I do n't believe in the slump of the market ,"he said .The computer company would offer open syst
ems so that customers would not be tied to a specific vendor or architecture ,and it will also focus on tying computers together th
rough communications networks .Reuter 3

Enter q to exit the browser or hit enter to continue
```

Δεδομένου ότι επιλέξουμε ένα από τα αποτελέσματα, τότε αυτό θα εμφανιστεί στο terminal. Το τελευταίο βήμα είναι η επιλογή από τον χρήστη για να συνεχίσει την χρήση του browser ή να επιλέξει την έξοδο από τον browser με το q.

## Queries

Η εφαρμογή αυτή καλύπτει πολλούς διαφορετικούς τύπους query για την επιστροφή των επιθυμητών αποτελεσμάτων.

### 1. Free-text Query

```
(base) stamatosorfanos@pop-os:~/git/SearchEngine$ cd /home/stamatosorfanos/git/SearchEngine ; /usr/bin/env /usr/lib/jvm/java-14-openjdk-amd64/bin/java -Xmx16G -XX:+ShowCodeDetailsInExceptionMessages -Dfile.encoding=UTF-8 -cp /home/stamatosorfanos/.config/Cod
e/User/workspaceStorage/1f103e0fbc3af7f96b6b6aa0a13ab204/redhat.java/jdt_ws/SearchEngine_6388bb4e/bin Application

Welcome to the browser
Please enter the query:
computer systems terminal sale
1./home/stamatosorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-000x9.txt

Please type the number of the page you would like to visit or 'c' to continue
```

### 2. Αυστηρό Boolean Query

```
(base) stamatosorfanos@pop-os:~/git/SearchEngine$ cd /home/stamatosorfanos/git/SearchEngine ; /usr/bin/env /usr/lib/jvm/java-14-openjdk-amd64/bin/java -Xmx16G -XX:+ShowCodeDetailsInExceptionMessages -Dfile.encoding=UTF-8 -cp /home/stamatosorfanos/.config/Cod
e/User/workspaceStorage/1f103e0fbc3af7f96b6b6aa0a13ab204/redhat.java/jdt_ws/SearchEngine_6388bb4e/bin Application

Welcome to the browser
Please enter the query:
computer AND systems AND terminal AND sale
1./home/stamatosorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-000x9.txt
```

### 3. Μίξη Free-Text και Boolean Query

```
(base) stamatosorfanos@pop-os:~/git/SearchEngine$ cd /home/stamatosorfanos/git/SearchEngine ; /usr/bin/env /usr/lib/jvm/java-14-openjdk-amd64/bin/java -Xmx16G -XX:+ShowCodeDetailsInExceptionMessages -Dfile.encoding=UTF-8 -cp /home/stamatosorfanos/.config/Cod
e/User/workspaceStorage/1f103e0fbc3af7f96b6b6aa0a13ab204/redhat.java/jdt_ws/SearchEngine_6388bb4e/bin Application

Welcome to the browser
Please enter the query:
computer systems AND terminal sale
1./home/stamatosorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-000x9.txt
```

```
(base) stamatosorfanos@pop-os:~/git/SearchEngine$ cd /home/stamatosorfanos/git/SearchEngine ; /usr/bin/env /usr/lib/jvm/java-14-openjdk-amd64/bin/java -Xmx16G -XX:+ShowCodeDetailsInExceptionMessages -Dfile.encoding=UTF-8 -cp /home/stamatosorfanos/.config/Cod
e/User/workspaceStorage/1f103e0fbc3af7f96b6b6aa0a13ab204/redhat.java/jdt_ws/SearchEngine_6388bb4e/bin Application

Welcome to the browser
Please enter the query:
computer systems terminal AND sale
1./home/stamatosorfanos/git/SearchEngine/Reuters8Kdata/data/acq/reut2-000x9.txt
```

Οι παραπάνω διαφορετικοί τρόποι κατασκευής ενός query ισχύουν και για τα query με OR operators με την διαφορά ότι συνεχόμενα terms θεωρούνται ως ένα AND query. Για παράδειγμα το query “family car OR family van”, θα επιστρέψει αποτελέσματα για family car ή family van καθώς θεωρούμε ότι το free-text κομμάτι καλύπτεται καλύτερα από το AND operator. Φυσικά με τον ίδιο τρόπο λειτουργούν και τα NOT queries με την μόνη διαφορά ο χρόνος εκτέλεσης τους είναι μεγαλύτερος από τα υπόλοιπα queries.