Μέρος 1.2.3

Για την δεύτερη φάση της προγραμματιστικής εργασίας χρησιμοποίησα τον κώδικα της WordNetLuceneDemo.java που δόθηκε και από κομμάτια του κώδικα από τον Searcher την πρώτη φάση της εργασίας .

Στο WordNet.java περνάω τα queries και όπως και στην φάση της εργασίας σε πίνακα και μετά την ολοκλήρωση της διαδικασίας αποθηκεύω τα αποτελέσματα στα κατάλληλα αρχεία με ονομασία my2ndresultsxx.test κτλπ.

Όπου xx είναι 20,30,50.

Για την σύγκριση των αποτελεσμάτων χρησιμοποιώ την Trec_Eval με την εντολή που είχαμε χρησιμοποιήσει και στην πρώτη φάση της εργασίας.

Πχ. >trec_eval -q -M 50 -m num_rel_ret -m map qrels.txt my2ndresults50.test

Για την εργασία πρώτα πρέπει να τρέξω τον indexer αν δεν υπάρχει ήδη το ευρετήριο με analyzer τον EnglishAnalyzer. Στη συνέχεια αφαιρώ από τα wn_s.pl τα ρήματα και τα αποθηκεύω και σε νέο αρχείο με όνομα wn_s2.pl, ελέγχω πρώτα αν υπάρχει το αρχείο αλλιώς το δημιουργώ. Το αρχείο θα πρέπει να υπάρχει στο φάκελο src, Με το παραδομένο zip, το αρχείο περιέχεται στο φάκελο. Αν θέλετε να το ξαναδημιουργήσετε απλά διαγράψτε το και ξαναπεράστε το στο φάκελο src.

Αφαιρώντας τα ρήματα όπως πρότεινε η καθηγήτρια βιντεοσκόπηση στην βελτιώνω το map από τα αποτελέσματα που έδινε από τον κώδικα του εργαστηρίου κατά 11,num_rel_ret στα 50 ενδεικτικά.

num_rel_ret	Q01	11
map	Q01	0.3875
num_rel_ret	Q02	3
map	Q02	0.0663
num_rel_ret	Q03	14
map	Q03	0.5745
num_rel_ret	Q04	4
map	Q04	0.0653
num_rel_ret	Q05	13
map	Q05	0.2619
num_rel_ret	Q06	1
map	Q06	0.0011
num_rel_ret	Q07	10
map	Q07	0.1747
num_rel_ret	Q08	10
map	Q08	0.4789
num_rel_ret	Q09	2
map	Q09	0.0273
num_rel_ret	Q10	2
map	Q10	0.0181
num_rel_ret	all	70
map	all	0.2056

num_rel_ret	Q01	15
map	Q01	0.6351
num_rel_ret	Q02	3
map	Q02	0.0663
num_rel_ret	Q03	14
map	Q03	0.5745
num_rel_ret	Q04	4
map	Q04	0.0653
num_rel_ret	Q05	13
map	Q05	0.2619
num_rel_ret	Q06	3
map	Q06	0.0250
num_rel_ret	Q07	10
map	Q07	0.1807
num_rel_ret	Q08	12
map	Q08	0.7056
num_rel_ret	Q09	5
map	Q09	0.0981
num_rel_ret	Q10	2
map	Q10	0.0181
num_rel_ret	all	81
map	all	0.2631

Επίσης παρακολουθώντας τα σχόλια των συμφοιτητών μου, άλλαξα κάποια φίλτρα στον Analyzer με αποτέλεσμα να μείνω με τα παρακάτω.

```
CustomAnalyzer.Builder builder = CustomAnalyzer.builder()

        .addTokenFilter(LowerCaseFilterFactory.class)
        .addTokenFilter(StopFilterFactory.class)
        .addTokenFilter(WordDelimiterGraphFilterFactory.class)
        .addTokenFilter(StandardFilterFactory.class)
        .addTokenFilter(EnglishPossessiveFilterFactory.class)
        .addTokenFilter(PorterStemFilterFactory.class)
        .addTokenFilter(SynonymGraphFilterFactory.class, sffargs)
        .addTokenFilter(RemoveDuplicatesTokenFilterFactory.class)
        .withTokenizer(StandardTokenizerFactory.class);
```

Αυτά δίνουν σαν αποτέλεσμα.

```
num_rel_ret             Q01     15
map                     Q01     0.6351
num_rel_ret             Q02     3
map                     Q02     0.0663
num_rel_ret             Q03     14
map                     Q03     0.5745
num_rel_ret             Q04     4
map                     Q04     0.0653
num_rel_ret             Q05     13
map                     Q05     0.2619
num_rel_ret             Q06     3
map                     Q06     0.0250
num_rel_ret             Q07     12
map                     Q07     0.2380
num_rel_ret             Q08     12
map                     Q08     0.7056
num_rel_ret             Q09     5
map                     Q09     0.0981
num_rel_ret             Q10     2
map                     Q10     0.0181
num_rel_ret             all     83
map                     all     0.2688
```

Συγκεντρωτικά τα βέλτιστα αποτελέσματα δίνονται στο excel, (δεν χωράνε στην σελιδα)

Θεωρω πως για την περαιτέρω βελτίωση τους και αντίστοιχα αυτό που μείωνει την απόδοση της εργασίας είναι το WordNet καθώς τα συνώνυμα δεν είναι πάντα παρόμοια, η έλλειψη δυνατότητας επιλογής συνωνύμων πχ, υπώνυμα όπως είδαμε στην βιντεοδιάλεξη.

Τέλος, δυσκολεύει την αποτελεσματική επιλογή εγγράφου με τα σύμβολα που εισάγει στο query το WordNet, όπως «(,),+» τα οποία δεν μπόρεσα να τα αφαιρέσω ούτε από το query αλλά ούτε εισάγοντάς επιπλέον filters και parameters.

Βιβλιογραφία Ενδεικτικά :

Eclass.aueb.gr

https://lucene.apache.org/core/6_4_2/analyzers-common/org/apache/lucene/analysis/custom/CustomAnalyzer.html

https://lucene.apache.org/core/5_5_5/analyzers-common/org/apache/lucene/analysis/custom/CustomAnalyzer.Builder.html

https://stackoverflow.com/questions/51989720/lucene-net-4-8-add-multiple-filters-to-custom-analyzer

https://lucene.apache.org/core/7_5_0/analyzers-common/org/apache/lucene/analysis/pattern/PatternReplaceCharFilterFactory.html