http://www.aueb.gr/images/logo_gr.gif



Συστήματα ανάκτης πληροφοριών

2012-2013

Παπουτσάκης Γιώργος 3040246

# 1η Φάση

Η λύση της πρώτης φάσης περιλαμβάνεται στο αρχείο Medlars\_collection.java. H συνάρτηση indexing() δημιουργεί το ευρετήριο και η main() μέθοδος του αρχείου διαβάζει το ευρετήριο αυτό καθώς και τα ερωτήματα από το αρχείο query.text και δημιουργεί το αρχείο results.text. Επίσης δημιουργείται το αρχείο qrels2.text καθώς το αρχικό δεν ήταν στη σωστή μορφή.Στο παραδοτέο αρχείο συμπεριλαμβάνονται όλα τα προαναφερθέντα αρχεία καθώς και το αρχείο trec\_results.text με τα πλήρη αποτελέσματα του trec\_eval ένα μικρό κομμάτι των οποίων φαίνεται παρακάτω .

Queryid (Num): 30

Total number of documents over all queries

Retrieved: 4812

Relevant: 696

Rel\_ret: 530

Interpolated Recall - Precision Averages:

at 0.00 0.3235

at 0.10 0.2897

at 0.20 0.2555

at 0.30 0.2371

at 0.40 0.2296

at 0.50 0.1784

at 0.60 0.1695

at 0.70 0.1626

at 0.80 0.1154

at 0.90 0.0702

at 1.00 0.0122

Average precision (non-interpolated) for all rel docs(averaged over queries) 0.1430

Precision:

At 5 docs: 0.1000

At 10 docs: 0.1333

At 15 docs: 0.1267

At 20 docs: 0.1500

At 30 docs: 0.1400

At 100 docs: 0.1160

At 200 docs: 0.0883

At 500 docs: 0.0353

At 1000 docs: 0.0177

R-Precision (precision after R (= num\_rel for a query) docs retrieved):

Exact: 0.1334

# 2η Φάση

Η λύση της δεύτερης φάσης περιλαμβάνεται στο αρχείο KnowledgeBase.java και Term.java. Οι συναρτήσεις parseTerms() και getTagValue() διαβάσουν το xml αρχείο και παράγουν μία λίστα από όρους.H μέθοδος indexing() χρησιμοποιώντας τις προηγούμενες δημιουργεί το ευρετήριο και η μέθοδος findSynonyms() επιστρέφει μια λίστα με τα συνώνυμα των όρων που είναι πιο κοντά στο ερώτημα που δέχεται ως είσοδο. Τέλος η επαύξηση του ερωτήματος γίνετε στο Medlars\_collection.java στο στάδιο που διαβάζεται το ερώτημα από το αρχείο χρησιμοποιώντας τη μέθοδο findSynonyms().

# 3η Φάση

Η λύση της τρίτης φάσης περιλαμβάνεται στο αρχείο Medlars\_collection.java στη συνάρτηση main() όπου και παράγονται τα αρχεία resultsexpand.text και resultsexpand10.text και με τη χρήση του trec\_eval τα αρχεία trec\_results\_expanded\_k5.text και trec\_results\_expanded\_k10.text που περιέχουν τα αποτελέσματα του trec\_eval στα ίδια ερωτήματα μόνο που στο 1ο το κάθε ερώτημα έχει επαυξηθεί με τα 5 πρώτα συνώνυμα του και το δεύτερο με τα 10 πρώτα. Τα αρχεία αυτά περιλαμβάνονται στο παραδοτέο αρχείο παρόλα αυτά ένα μικρό απόσπασμα τους φαίνεται παρακάτω.

trec\_results\_expanded\_k5

Queryid (Num): 30

Total number of documents over all queries

Retrieved: 5712

Relevant: 696

Rel\_ret: 546

Interpolated Recall - Precision Averages:

at 0.00 0.2512

at 0.10 0.2335

at 0.20 0.1991

at 0.30 0.1814

at 0.40 0.1763

at 0.50 0.1564

at 0.60 0.1346

at 0.70 0.1235

at 0.80 0.0805

at 0.90 0.0521

at 1.00 0.0098

Average precision (non-interpolated) for all rel docs(averaged over queries)

0.1063

Precision:

At 5 docs: 0.0600

At 10 docs: 0.0800

At 15 docs: 0.0689

At 20 docs: 0.0717

At 30 docs: 0.0733

At 100 docs: 0.0933

At 200 docs: 0.0910

At 500 docs: 0.0364

At 1000 docs: 0.0182

R-Precision (precision after R (= num\_rel for a query) docs retrieved):

Exact: 0.0886

trec\_results\_expanded\_k10

Queryid (Num): 30

Total number of documents over all queries

Retrieved: 5814

Relevant: 696

Rel\_ret: 549

Interpolated Recall - Precision Averages:

at 0.00 0.2552

at 0.10 0.2405

at 0.20 0.2013

at 0.30 0.1826

at 0.40 0.1735

at 0.50 0.1545

at 0.60 0.1432

at 0.70 0.1280

at 0.80 0.1035

at 0.90 0.0420

at 1.00 0.0101

Average precision (non-interpolated) for all rel docs(averaged over queries)

0.1089

Precision:

At 5 docs: 0.0533

At 10 docs: 0.0900

At 15 docs: 0.0844

At 20 docs: 0.0850

At 30 docs: 0.0822

At 100 docs: 0.0927

At 200 docs: 0.0915

At 500 docs: 0.0366

At 1000 docs: 0.0183

R-Precision (precision after R (= num\_rel for a query) docs retrieved):

Exact: 0.0945

# 4η Φάση









