Συστήματα Ανάκτησης Πληροφοριών

Φάση 3 – Πιθανοτικό και Γλωσσικό Μοντέλο Ανάκτησης

Στυλιανή Δούκα – p3170042

Βασίλειος Μπάλλας – p170115

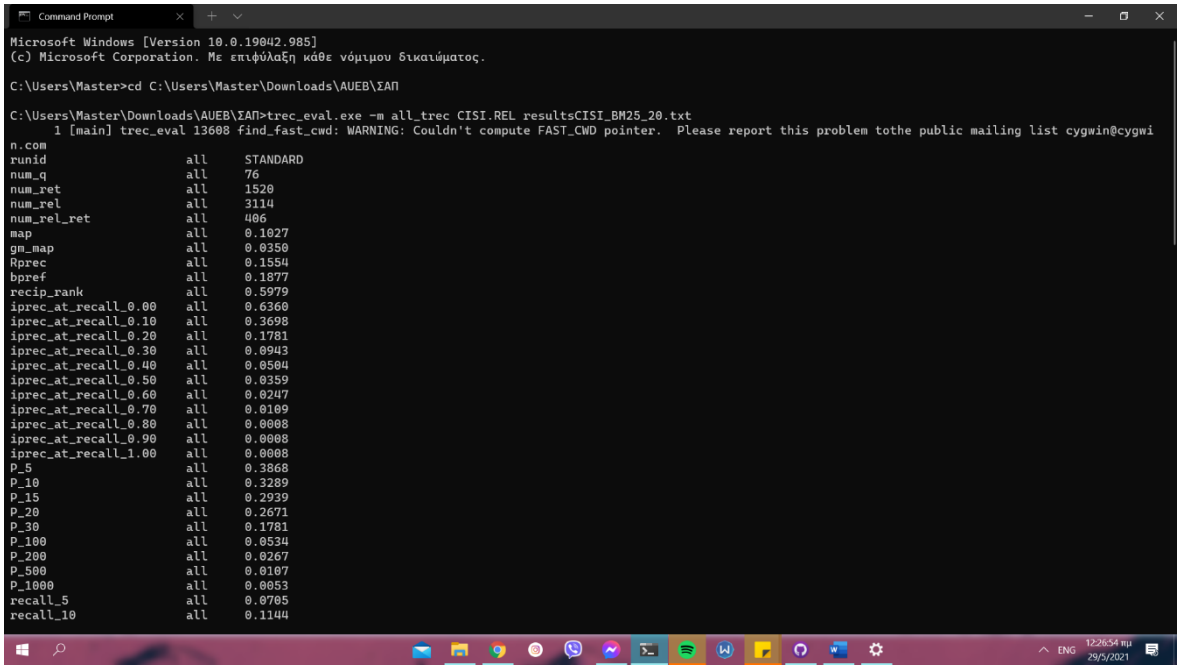
Περιγραφή Υλοποίησης

Για τις ανάγκες αυτής της Φάσης της εργασίας αλλάξαμε τον τύπο της κλάσης Ομοιότητας πρωτα σε BM25Similarity και έπειτα σε LMJelinekMercerSimilarity.

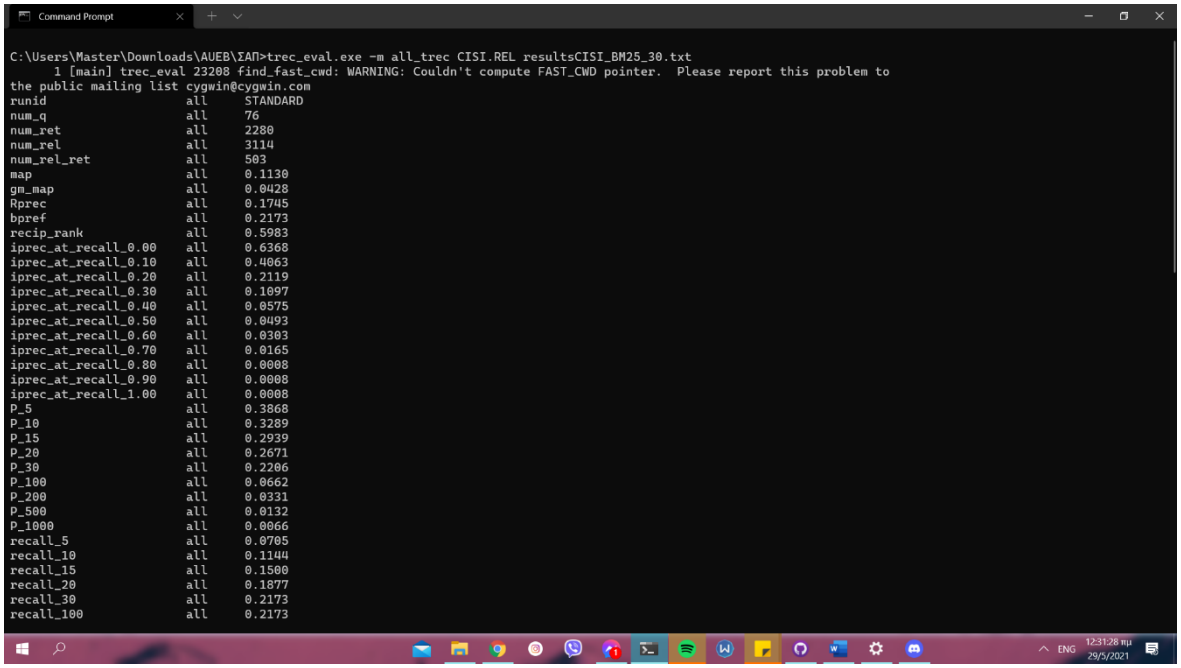
```
// Access the index using indexReaderFSDirectory.open(Paths.get(index)
IndexReader indexReader = DirectoryReader.open(FSDirectory.open(Paths.
IndexSearcher indexSearcher = new IndexSearcher(indexReader);
indexSearcher.setSimilarity(new BM25Similarity());
indexSearcher.setSimilarity(new LMJelinekMercerSimilarity(0.7f));
```

Στη συνέχεια περάσαμε τα αποτελέσματα από το Trec_eval:

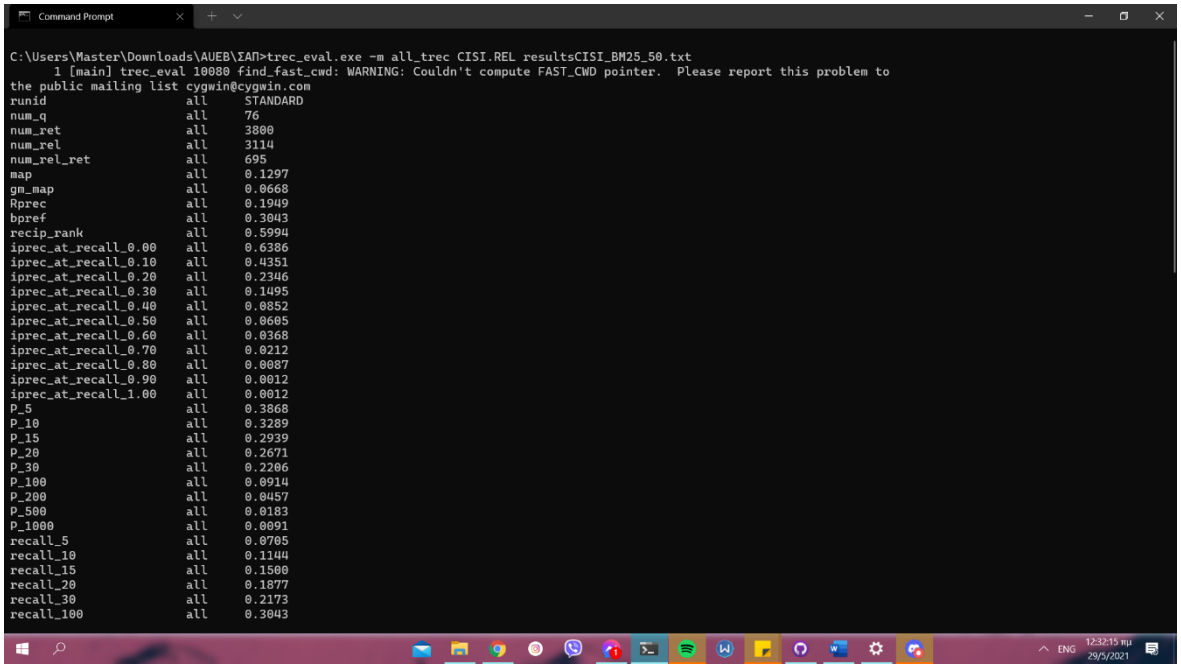
Εκτελούμε trec_eval.exe -m all_trec CISI.REL resultsCISI_BM25_20.txt στα 20 πιο σχετικά κείμενα



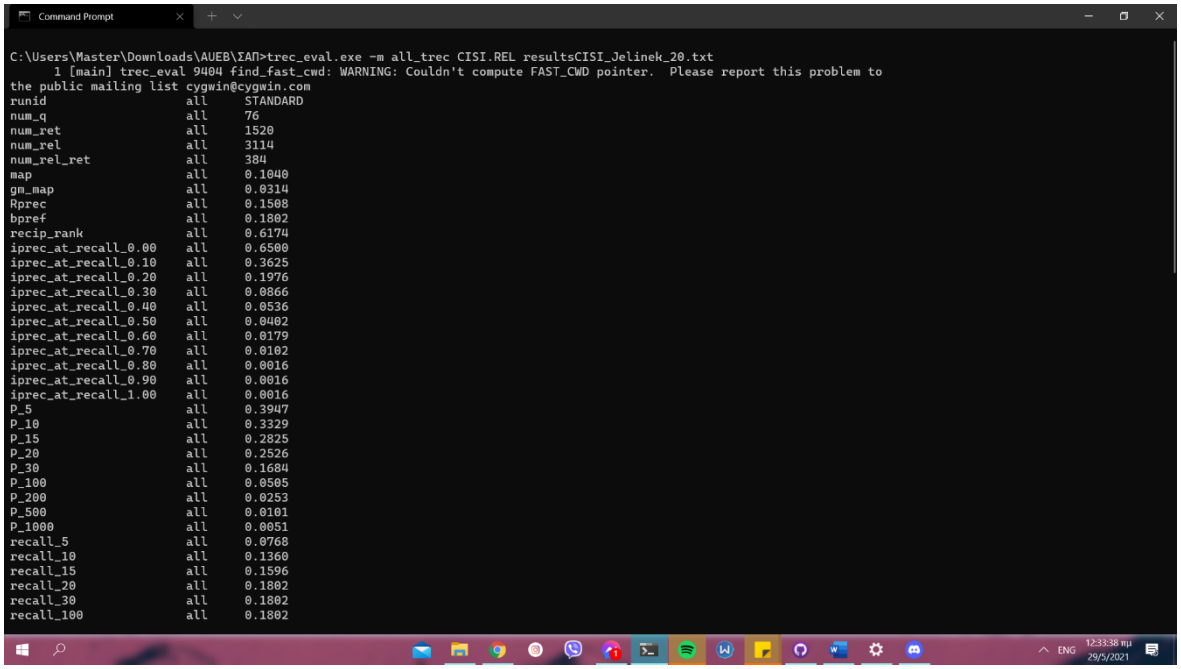
Εκτελούμε trec_eval.exe -m all_trec CISI.REL resultsCISI_BM25_30.txt στα 30 πιο σχετικά κείμενα



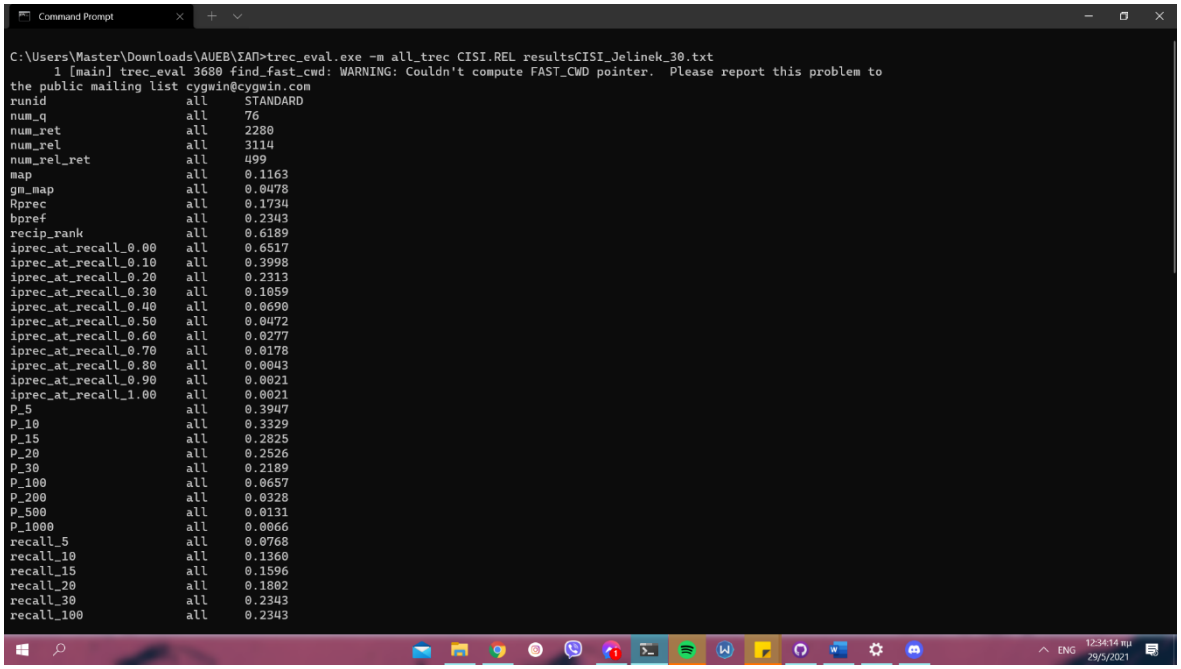
Εκτελούμε trec_eval.exe -m all_trec CISI.REL resultsCISI_BM25_50.txt στα 50 πιο σχετικά κείμενα



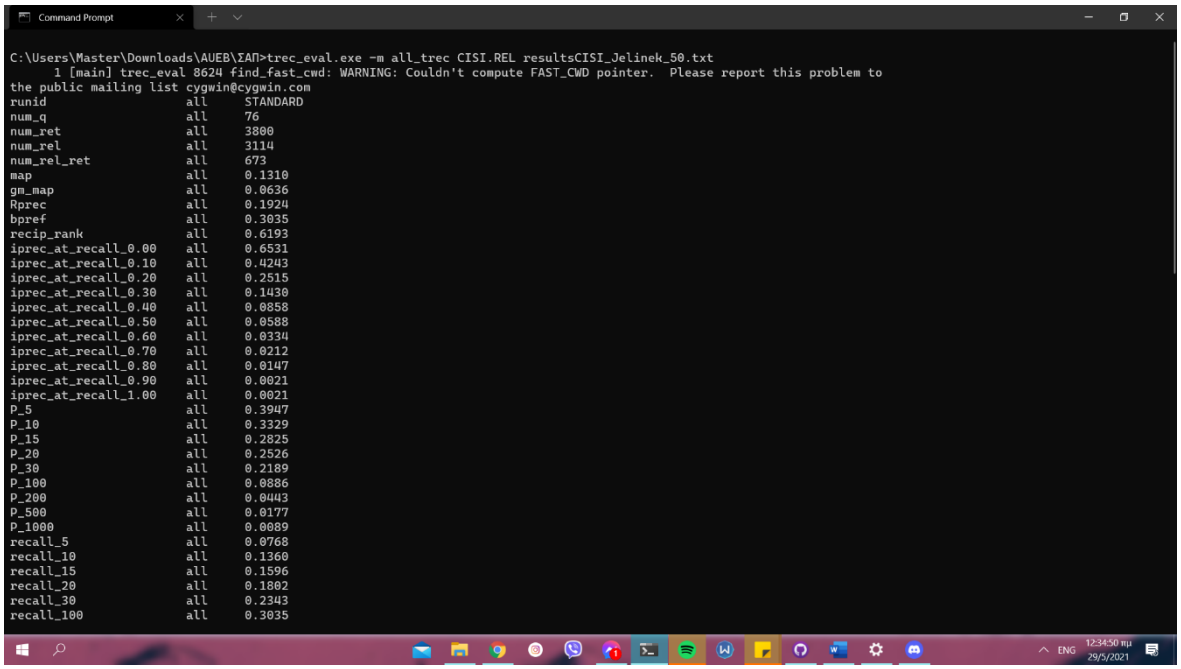
Εκτελούμε trec_eval.exe -m all_trec CISI.REL resultsCISI_Jelinek_20.txt στα 20 πιο σχετικά κείμενα



Εκτελούμε trec_eval.exe -m all_trec CISI.REL resultsCISI_Jelinek_30.txt στα 30 πιο σχετικά κείμενα



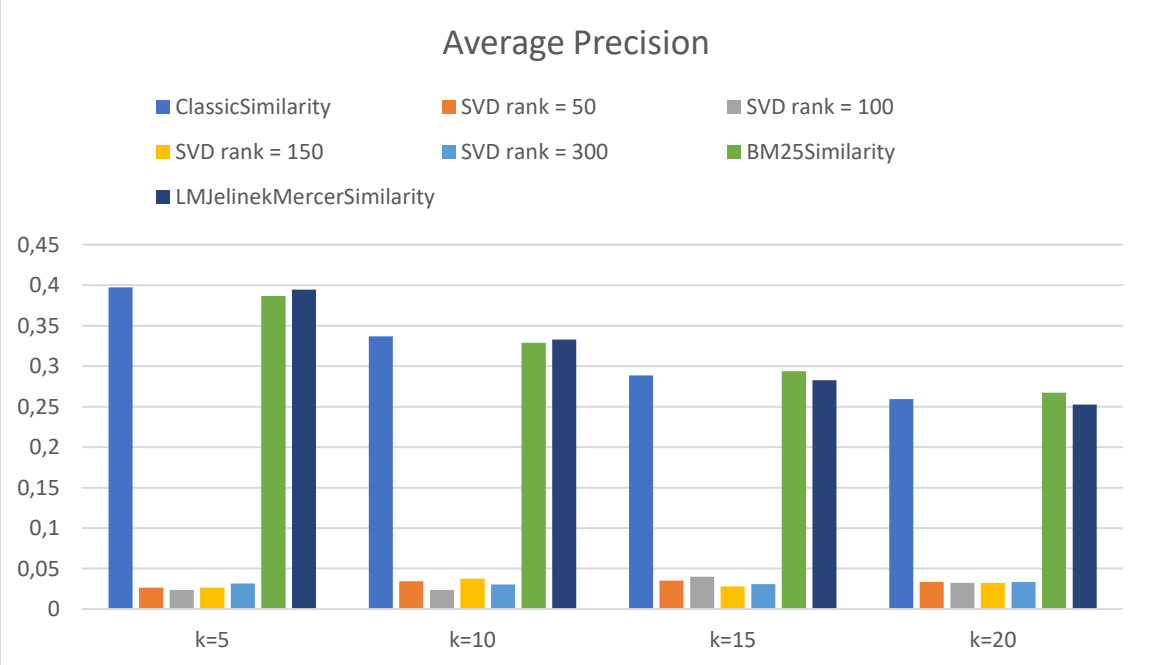
Εκτελούμε trec_eval.exe -m all_trec CISI.REL resultsCISI_Jelinek_50.txt στα 50 πιο σχετικά κείμενα



Σύγκριση με προηγούμενες Φάσεις

Average Precision

	k=5	k=10	k=15	k=20
ClassicSimilarity	0,3974	0,3368	0,2886	0,2592
SVD rank = 50	0,0263	0,0342	0,0351	0,0336
SVD rank = 100	0,0237	0,0237	0,0398	0,0322
SVD rank = 150	0,0263	0,0376	0,0281	0,0322
SVD rank = 300	0,0316	0,0303	0,0307	0,0336
BM25Similarity	0,3868	0,3289	0,2939	0,2671
LMJelinekMercerSimilarity	0,3947	0,3329	0,2825	0,2526

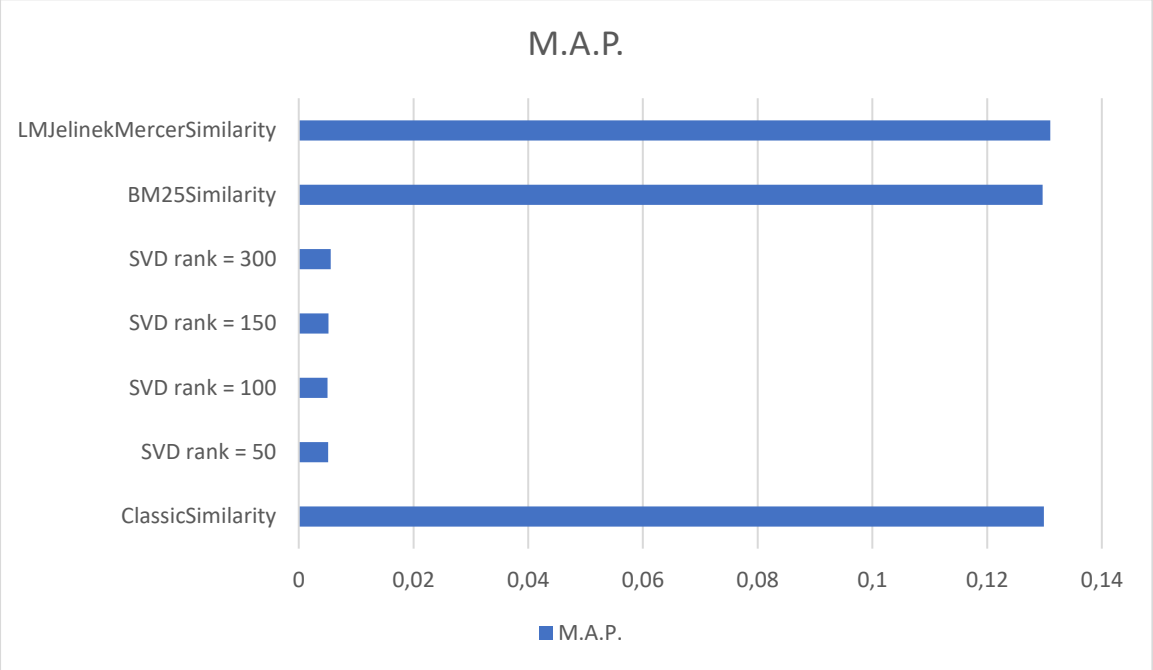


Παρατηρούμε ότι ο SVD βγαζει πολύ χειρότερα αποτελέσματα από τις κλάσεις ομοιότητας. Οι κλάσεις που χρησιμοποιήσαμε στη Φάση 3 έχουν παρόμοια αποτελέσματα με την ClassicSimilarity που χρησιμοποιήσαμε στην Φαση 1.

Παρόμοια είναι και η κατάσταση στο M.A.P.:

Mean Average Precision

	M.A.P.
ClassicSimilarity	0,1299
SVD rank = 50	0,0051
SVD rank = 100	0,0050
SVD rank = 150	0,0052
SVD rank = 300	0,0056
BM25Similarity	0,1297
LMJelinekMercerSimilarity	0,1310



Πηγές:

https://lucene,apache.org/core/8_0_0/core/org/apache/lucene/search/similarities/LMJelinekMercerSimilarity.html