Συστήματα Ανάκτησης Πληροφοριών

Φάση 5 ‐ Ανάκτηση χρησιμοποιώντας πολλαπλές συναρτήσεις ομοιότητας

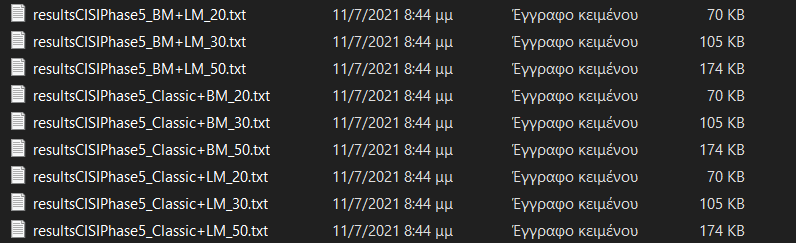
Στυλιανή Δούκα – p3170042

Βασίλειος Μπάλλας – p3170115

**Περιγραφή Υλοποίησης**

Για τις ανάγκες αυτής της Φάσης της εργασίας επιστρέψαμε στο project της 3ης Φάσης και το αξιοποιήσαμε εφαρμόζοντας συνδυαστικά τις συναρτήσεις ομοιότητας που είχαμε χρησιμοποιήσει στις φάσεις 1 και 3. Παρακάτω αναλύουμε τους διάφορους συνδυασμούς που δοκιμάστηκαν.

Αποθηκεύουμε τα αποτελέσματα σε αρχεία .txt



Στη συνέχεια περάσαμε τα αποτελέσματα από το Trec\_eval

**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 |
| WordEmbeddingsSimilarity | 0,0184 | 0,0250 | 0,0246 | 0,0283 |
| PretrainedEmbeddingsWiki | 0,1711 | 0,1408 | 0,1333 | 0,1204 |
| Classic+BM25 | 0,3816 | 0,3342 | 0,2956 | 0,2651 |
| Classic+LMJelinekMercer | 0,4053 | 0,3316 | 0,2851 | 0,2572 |
| BM25+LMJelinekMercer | 0,3868 | 0,3342 | 0,2912 | 0,2618 |

Είναι φανερό ότι ο συνδυασμός των συναρτήσεων ομοιότητας αποδείχθηκε πιο αποδοτικός από κάθε μία ξεχωριστά.

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 |
| WordEmbeddingsSimilarity | 0,0039 |
| PretrainedEmbeddingsWiki | 0,0370 |
| Classic+BM25 | 0,1309 |
| Classic+LMJelinekMercer | 0,1312 |
| BM25+LMJelinekMercer | 0,1305 |

Πηγές:

<https://lucene.apache.org/core/8_0_0/core/org/apache/lucene/search/similarities/MultiSimilarity.html>