



ΣΥΣΤΗΜΑΤΑ ΑΝΑΚΤΗΣΗΣ ΠΛΗΡΟΦΟΡΙΩΝ

4ο Παραδοτέο



ΜΠΙΛΙΑΣ ΓΕΩΡΓΙΟΣ 3200278
ΙΩΑΝΝΗΣ ΠΑΤΟΥΧΑΣ 3200149

Για τη φάση 4 χρησιμοποιήσαμε σχεδόν τον ίδιο κώδικα με την πρώτη φάση, εκτός από τη χρήση υλοποίησης του multisimilarity , στην phase4.java :

```
MultiSimilarity similarity = new MultiSimilarity(new Similarity[]{  
    new BM25Similarity(3, 1),  
    new LMJelinekMercerSimilarity(0.9F),  
});
```

Και στην searcher4.java παρόμοια (γραμμές 36-39)

Η επιλογή των παραμέτρων έγινε με δοκιμές συνδυασμών βλέποντας κάθε φορά τα αποτελέσματα.

Τα αποτελέσματα του trec eval του καλύτερου συνδυασμού ήταν τα εξής :

<u>runid</u>	all <u>myIRmethod</u>
<u>num_q</u>	all 10
<u>num_ret</u>	all 500
<u>num_rel</u>	all 152
<u>num_rel_ret</u>	all 97
<u>map</u>	all 0.3767
<u>gm_map</u>	all 0.2728
<u>Rprec</u>	all 0.4066
<u>bpref</u>	all 0.6396
<u>recip_rank</u>	all 0.7267
<u>iprec_at_recall_0.00</u>	all 0.7912
<u>iprec_at_recall_0.10</u>	all 0.6672
<u>iprec_at_recall_0.20</u>	all 0.6117
<u>iprec_at_recall_0.30</u>	all 0.5342
<u>iprec_at_recall_0.40</u>	all 0.4642
<u>iprec_at_recall_0.50</u>	all 0.3292
<u>iprec_at_recall_0.60</u>	all 0.2949
<u>iprec_at_recall_0.70</u>	all 0.2549
<u>iprec_at_recall_0.80</u>	all 0.2165
<u>iprec_at_recall_0.90</u>	all 0.1260
<u>iprec_at_recall_1.00</u>	all 0.0842
<u>P_5</u>	all 0.5200
<u>P_10</u>	all 0.4200
<u>P_15</u>	all 0.3933
<u>P_20</u>	all 0.3350

Παρατηρούμε πως είναι τα καλύτερα από όλες τις φάσεις με ελάχιστη βελτίωση από το μέγιστο \max της φάσης 2.