Συστήματα Ανάκτησης Πληροφοριών

2ο Παραδοτέο

MAY 25 2024

ΜΠΙΛΙΑΣ ΓΕΩΡΓΙΟΣ 3200278 ΙΩΑΝΝΗΣ ΠΑΤΟΥΧΑΣ 3200149

ΠΕΡΙΕΧΟΜΕΝΑ

LMJelinekMercerSimilarity

Για τη χρηση του **LMJelinekMercerSimilarity** αντικαταστήσαμε τις 2 γραμμές κώδικα που αφορούσαν το ClassicSimilarity ως εξής :

```
LMJelinekMercerSimilarity similarity = new LMJelinekMercerSimilarity(0.9f);
```

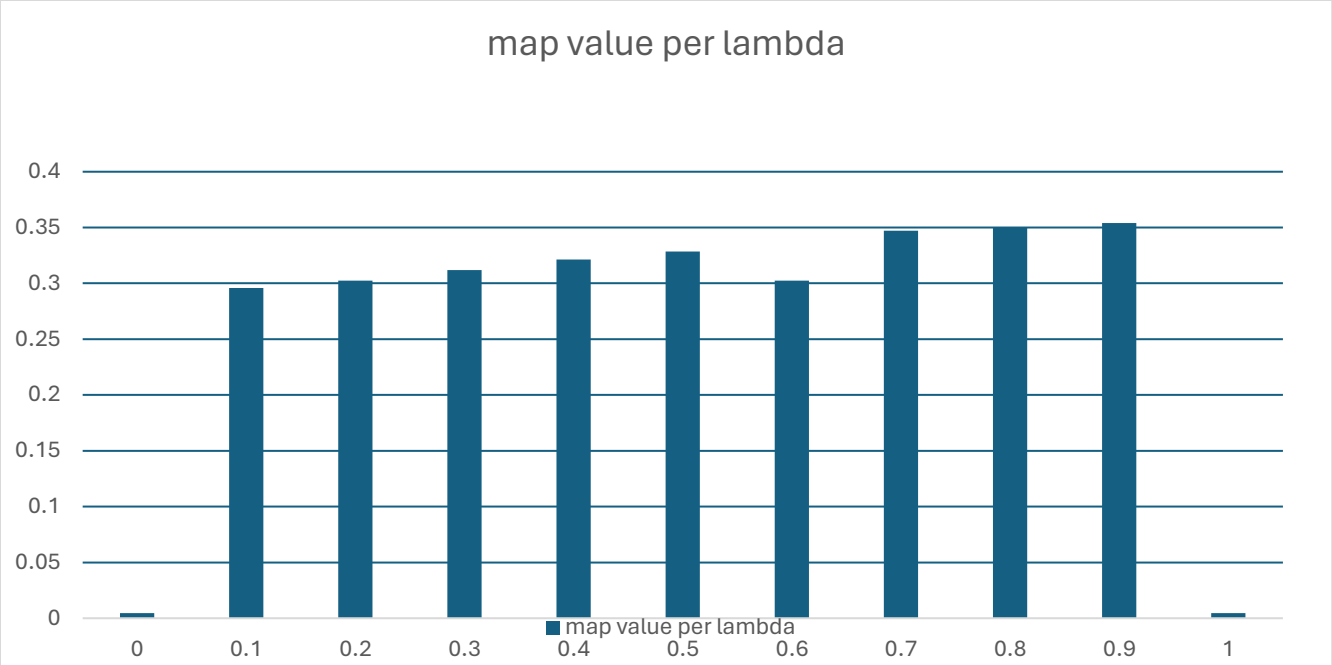
Στη γραμμη 34 του αρχείου searcher.java

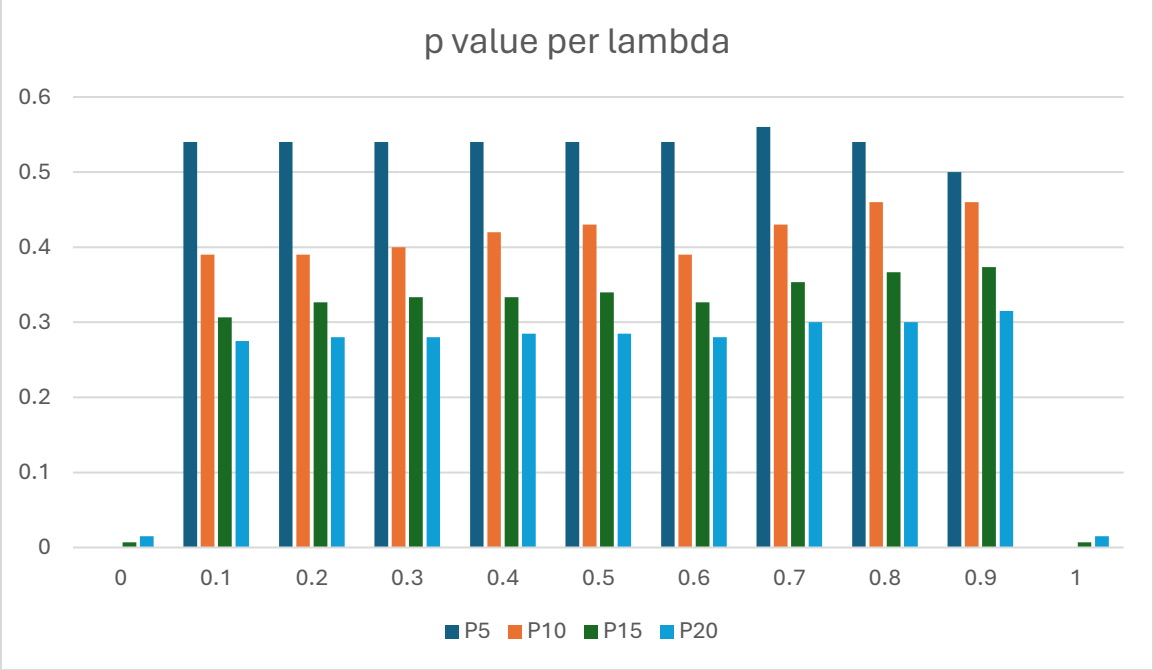
```
indexSearcher.setSimilarity(new LMJelinekMercerSimilarity(0.9f));
```

Στη γραμμη 40 του αρχείου FirstPhase.java

Δοκιμάσαμε αρκετές τιμές για την παράμετρο lambda όπως φαίνεται και στον παρακάτω πίνακα και γράφημα καταλήξαμε στο συμπέρασμα πως για **λ = 0.9** παρατηρούμε τα καλύτερο map αλλά όχι καλύτερο από όταν χρησιμοποιηούσαμε το ClassicSimilarity

λ (lambda)	map	P5	P10	P15	P20
0	0.0047	0.0000	0.0000	0.0067	0.0150
0.1	0.2958	0.5400	0.3900	0.3067	0.2750
0.2	0.3025	0.5400	0.3900	0.3267	0.2800
0.3	0.3120	0.5400	0.4000	0.3333	0.2800
0.4	0.3214	0.5400	0.4200	0.3333	0.2850
0.5	0.3286	0.5400	0.4300	0.3400	0.2850
0.6	0.3025	0.5400	0.3900	0.3267	0.2800
0.7	0.3471	0.5600	0.4300	0.3533	0.3000
0.8	0.3504	0.5400	0.4600	0.3667	0.3000
0.9	0.3541	0.5000	0.4600	0.3733	0.3150
1	0.0047	0.0000	0.0000	0.0067	0.0150





Όλα τα αποτελέσματα του trec eval για **λ = 0.9** :

```
runid                   all myIRmethod
num_q                   all 10
num_ret                 all 500
num_rel                 all 152
num_rel_ret             all 94
map                     all 0.3541
gm_map                  all 0.2592
Rprec                   all 0.3879
bpref                   all 0.6219
recip_rank              all 0.7167
iprec_at_recall_0.00    all 0.7914
iprec_at_recall_0.10    all 0.6764
iprec_at_recall_0.20    all 0.6076
iprec_at_recall_0.30    all 0.4937
iprec_at_recall_0.40    all 0.4285
iprec_at_recall_0.50    all 0.3205
iprec_at_recall_0.60    all 0.2710
iprec_at_recall_0.70    all 0.2415
iprec_at_recall_0.80    all 0.1765
iprec_at_recall_0.90    all 0.0957
iprec_at_recall_1.00    all 0.0311
P_5                     all 0.5000
P_10                    all 0.4600
P_15                    all 0.3733
P_20                    all 0.3150
```

BM25Similarity

Για τη χρηση του **BM25Similarity** αντικαταστήσαμε τις 2 γραμμές κώδικα που αφορούσαν το ClassicSimilarity ως εξής :

```
BM25Similarity similarity = new BM25Similarity(3, 1);
```

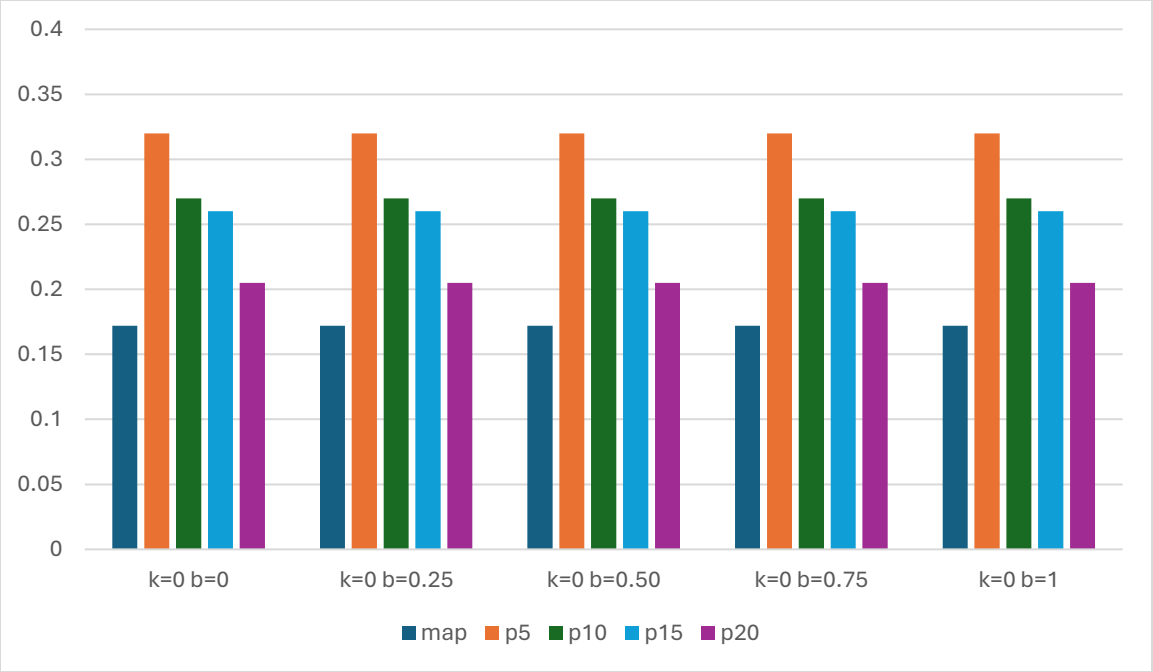
Στη γραμμη 34 του αρχείου searcher.java

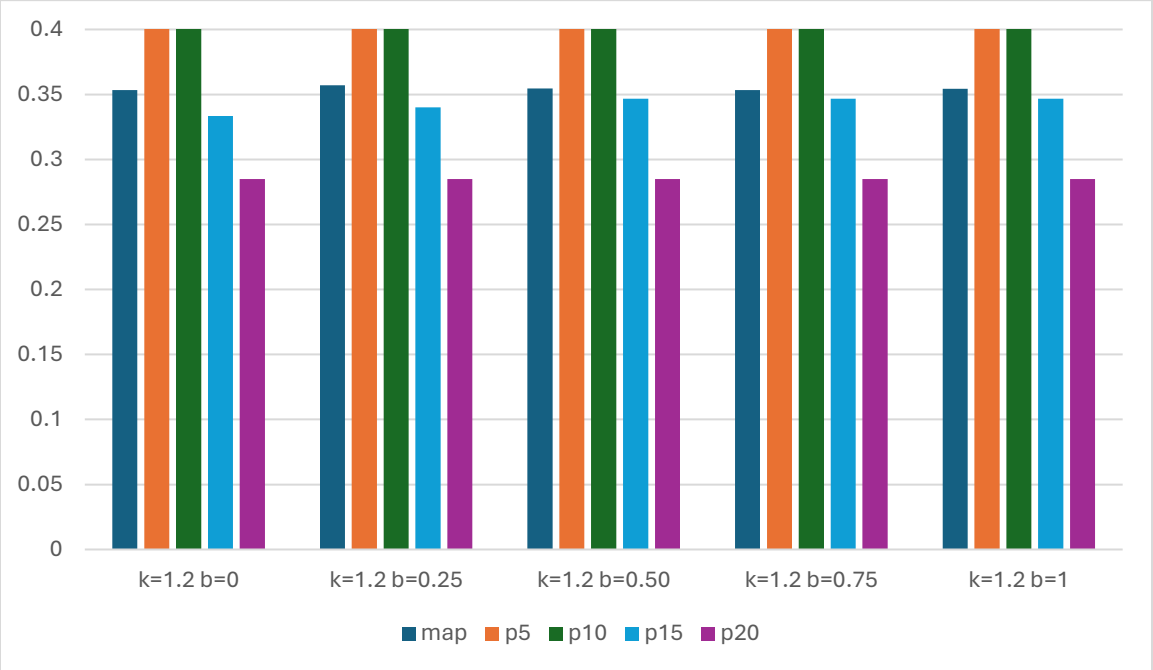
```
indexSearcher.setSimilarity(new BM25Similarity(3, 1));
```

Στη γραμμη 40 του αρχείου FirstPhase.java

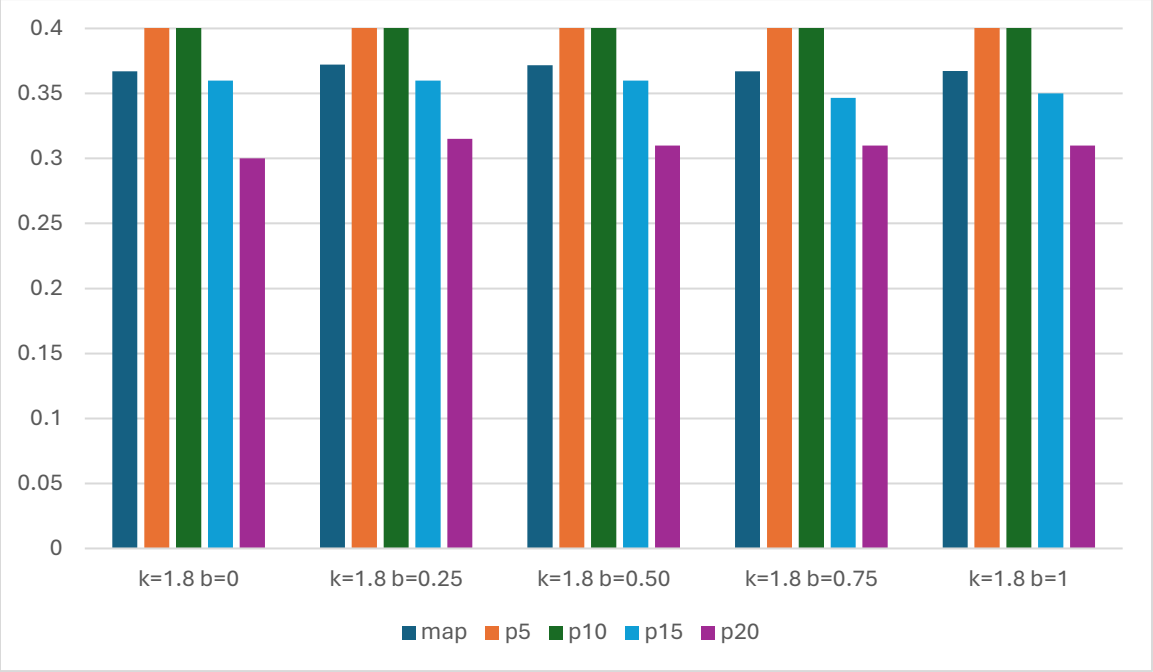
Δοκιμάσαμε αρκετές τιμές για τις παραμέτρους b ∈ (0,1) και k όπου k ≥ 0 και όχι ∞ , (όπως φαίνεται και στον παρακάτω πίνακα και γράφημα) και καταλήξαμε στο συμπέρασμα πως για k=3 και b=1 παρατηρούμε τα καλύτερο map

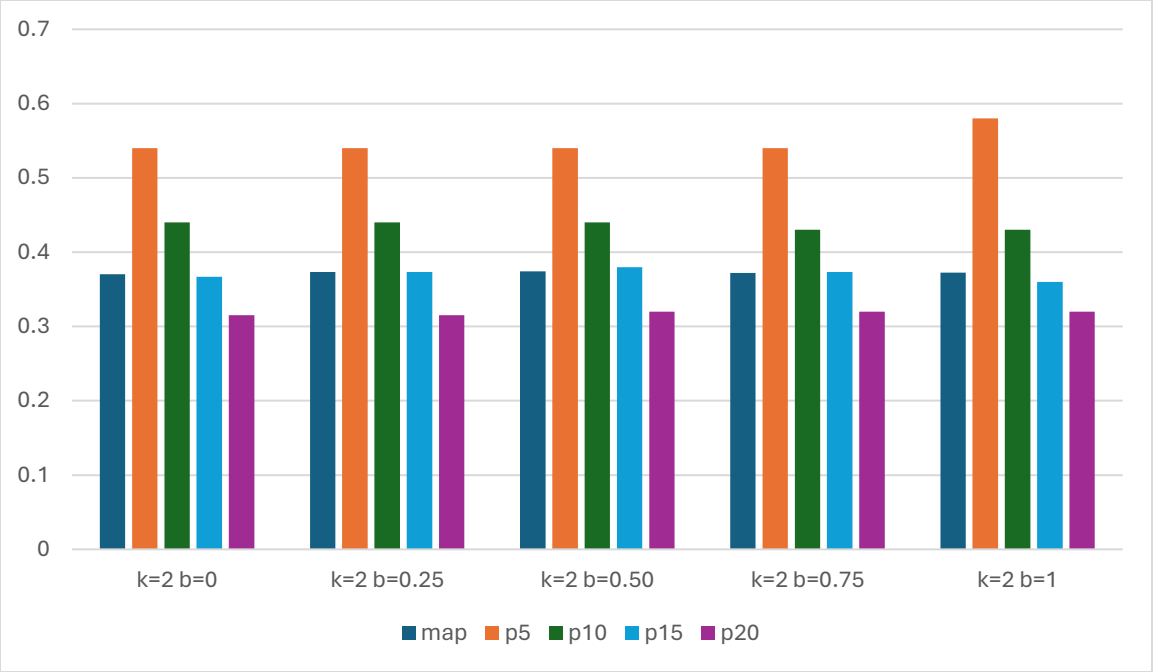
K b	0	0.25	0.50	0.75	1
0	map = 0.1721 P5 = 0.3200 P10 = 0.2700 P15 = 0.2600 P20 = 0.2050	map = 0.1721 p5 = 0.3200 p10 = 0.2700 p15 = 0.2600 p20 = 0.2050	map = 0.1721 p5 = 0.3200 p10 = 0.2700 p15 = 0.2600 p20 = 0.2050	0.1721 p5 = 0.3200 p10 = 0.2700 p15 = 0.2600 p20 = 0.2050	0.1721 p5 = 0.3200 p10 = 0.2700 p15 = 0.2600 p20 = 0.2050
1.2	map = 0.3533 p5 = 0.5600 p10 = 0.4400 p15 = 0.3333 p20 = 0.2850	map = 0.3571 p5 = 0.5800 p10 = 0.4400 p15 = 0.3400 p20 = 0.2850	map = 0.3546 p5 = 0.5800 p10 = 0.4300 p15 = 0.3467 p20 = 0.2850	map = 0.3534 p5 = 0.5800 p10 = 0.4400 p15 = 0.3467 p20 = 0.2850	map = 0.3544 p5 = 0.6000 p10 = 0.4400 p15 = 0.3467 p20 = 0.2850
1.5	map = 0.3628 p5 = 0.5400 p10 = 0.4400 p15 = 0.3467 p20 = 0.3050	map = 0.3646 p5 = 0.5600 p10 = 0.4400 p15 = 0.3400 p20 = 0.3050	map = 0.3645 p5 = 0.5600 p10 = 0.4500 p15 = 0.3400 p20 = 0.3100	map = 0.3618 p5 = 0.5600 p10 = 0.4300 p15 = 0.3400 p20 = 0.3050	map = 0.3654 p5 = 0.6000 p10 = 0.4300 p15 = 0.3400 p20 = 0.3000
1.8	map = 0.3671 p5 = 0.5400 p10 = 0.4400 p15 = 0.3600 p20 = 0.3050	map = 0.3721 p5 = 0.5400 p10 = 0.4400 p15 = 0.3600 p20 = 0.3150	map = 0.3717 p5 = 0.5400 p10 = 0.4400 p15 = 0.3600 p20 = 0.3100	map = 0.3671 p5 = 0.5600 p10 = 0.4300 p15 = 0.3467 p20 = 0.3150	map = 0.3672 p5 = 0.6000 p10 = 0.4300 p15 = 0.3533 p20 = 0.3100
2	map = 0.3703 p5 = 0.5400 p10 = 0.4400 p15 = 0.3667 p20 = 0.3150	map = 0.3732 p5 = 0.5400 p10 = 0.4400 p15 = 0.3733 p20 = 0.3150	map = 0.3741 p5 = 0.5400 p10 = 0.4400 p15 = 0.3800 p20 = 0.3200	map = 0.3721 p5 = 0.5400 p10 = 0.4300 p15 = 0.3733 p20 = 0.3200	map = 0.3725 p5 = 0.5800 p10 = 0.4300 p15 = 0.3600 p20 = 0.3200
3	map = 0.3753 p5 = 0.5200 p10 = 0.4300 p15 = 0.3867 p20 = 0.3350	map = 0.3736 p5 = 0.5200 p10 = 0.4300 p15 = 0.3867 p20 = 0.3350	map = 0.3735 p5 = 0.5000 p10 = 0.4300 p15 = 0.3867 p20 = 0.3350	map = 0.3749 p5 = 0.5200 p10 = 0.4300 p15 = 0.3867 p20 = 0.3400	map = 0.3760 p5 = 0.5400 p10 = 0.4200 p15 = 0.3933 p20 = 0.3200

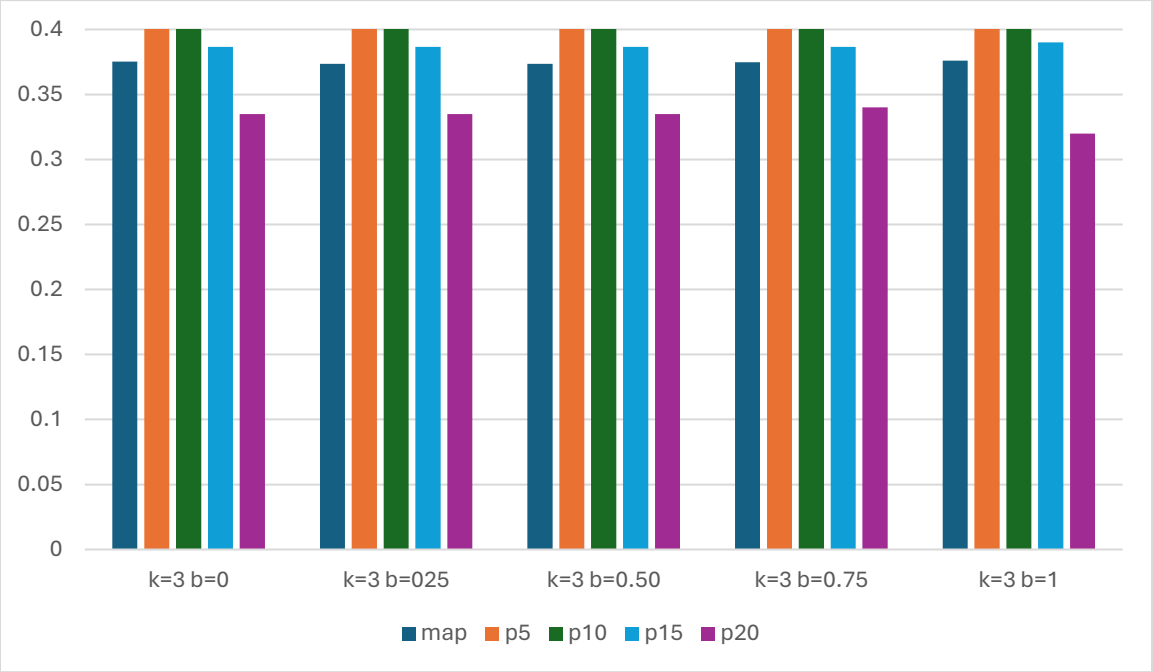

0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
k=0 b=0
k=0 b=0.25
k=0 b=0.50
k=0 b=0.75
k=0 b=1
map
p5
p10
p15
p20


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
k=1.2 b=0
k=1.2 b=0.25
k=1.2 b=0.50
k=1.2 b=0.75
k=1.2 b=1
map
p5
p10
p15
p20


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
k=1.5 b=0
k=1.5 b=0.25
k=1.5 b=0.50
k=1.5 b=0.75
k=1.5 b=1
map
p5
p10
p15
p20


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
k=1.8 b=0
k=1.8 b=0.25
k=1.8 b=0.50
k=1.8 b=0.75
k=1.8 b=1
map
p5
p10
p15
p20


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
k=2 b=0
k=2 b=0.25
k=2 b=0.50
k=2 b=0.75
k=2 b=1
map
p5
p10
p15
p20


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
k=3 b=0
k=3 b=025
k=3 b=0.50
k=3 b=0.75
k=3 b=1
map
p5
p10
p15
p20

Όλα τα αποτελέσματα του trec eval για **k=3 , b=1** :

```
runid                   all myIRmet
num_q                   all 10
num_ret                 all 500
num_rel                 all 152
num_rel_ret             all 95
map                     all 0.3760
gm_map                  all 0.2694
Rprec                   all 0.4001
bpref                   all 0.6301
recip_rank              all 0.7144
iprec_at_recall_0.00    all 0.7756
iprec_at_recall_0.10    all 0.6619
iprec_at_recall_0.20    all 0.6318
iprec_at_recall_0.30    all 0.5230
iprec_at_recall_0.40    all 0.4901
iprec_at_recall_0.50    all 0.3302
iprec_at_recall_0.60    all 0.2920
iprec_at_recall_0.70    all 0.2511
iprec_at_recall_0.80    all 0.1917
iprec_at_recall_0.90    all 0.1219
iprec_at_recall_1.00    all 0.0783
P_5                     all 0.5400
P_10                    all 0.4200
P_15                    all 0.3933
P_20                    all 0.3200
```