

Πανεπιστήμιο Κρήτης
Τμήμα Επιστήμης Υπολογιστών
ΗΥ463 Συστήματα Ανάκτησης Πληροφοριών
Εξάμηνο: Άνοιξη 2019

Γραπτή Αναφορά Έργου

Φαση Β΄

Στοιχεία Φοιτητών

Μέλος	1 ^ο
Ονοματwpώνυμο	Πρινιανάκης Αντώνης
ΑΜ	3833
Email	csd3833@csd.uoc.gr

Μέλος	2 ^ο
Ονοματwpώνυμο	Σαμαριτάκη Γεωργία
ΑΜ	3840
Email	csd3840@csd.uoc.gr

1 Περιεχόμενα

2	ΕΙΣΑΓΩΓΗ	2
3	ΜΕΤΡΗΣΕΙΣ.....	2
4	ΕΠΙΛΟΓΟΣ.....	4
5	ΑΝΑΦΟΡΕΣ	5

2 Εισαγωγή

Στην δευτερη φάση του προτζεκτ αυτού τρέξαμε στο σύστημα μας 30 topics που μας είχαν δοθεί. Δημιουργήσαμε queries από αυτά και μαζέψαμε τα αποτελέσματα για να τα αξιολογήσουμε. Στη συνέχεια, υπολογίσαμε τις μετρικές bpref, Asep, nDCG για να βγάλουμε συμπεράσματα.

3 Μετρήσεις

Αρχικά επιχειρήσαμε να παράγουμε τα queries από το description δημιουργώντας έτσι μεγαλύτερα και πιο αργά searches. Τα descriptions όμως έχουν λέξεις που δεν σχετίζονται με την αναζήτηση που θέλει να γίνει με αποτέλεσμα να θεωρούνται περισσότερα αρχεία ως σχετικά. Στην συνέχεια παράξαμε τα queries από το summary πράγμα που επίσπευσε την διαδικασία και έβγαλε γενικά καλύτερα αποτελέσματα.

Ο χρόνος εκτέλεσης για την παραγωγή αποτελεσμάτων και για την αξιολόγηση ήταν 45λεπτα.

Τα αποτελέσματα ήταν τα εξής:

Topic	Bpref	Asep	nDCG	Sum
1	0.647406611	0.482252688	0.724194644	1.85385394
2	0.909260098	0.480450015	0.543682023	1.93339214
3	0	0.121347994	0.370749449	0.49209744
4	0.875432526	0.721590801	0.86095243	2.45797576
5	0.312744141	0.217802899	0.38952401	0.92007105
6	0.830629528	0.425489426	0.592103278	1.84822223
7	0.827755178	0.819823808	0.970473258	2.61805224
8	0.90106824	0.482226139	0.616334725	1.9996291
9	0	0.222816075	0.502663166	0.72547924
10	0.588888889	0.718393202	0.825963249	2.13324534
11	0	0.055339883	0.163235074	0.21857496
12	0.294270833	0.140611133	0.305738579	0.74062055
13	0.446745562	0.274959828	0.402726334	1.12443172
14	0.7344	0.402997705	0.613373657	1.75077136
15	0.773511342	0.701713744	0.850481655	2.32570674
16	0	0	0	0
17	0.638035047	0.545102319	0.819483141	2.00262051
18	0.675428826	0.497589691	0.726217834	1.89923635
19	0.87015625	0.708358062	0.718196294	2.29671061
20	0.7727797	0.502814602	0.678760061	1.95435436
21	0.824446068	0.48913098	0.645057158	1.95863421
22	0	0.208731451	0.442620382	0.65135183
23	0	0.109523227	0.305476738	0.41499996
24	0.099112426	0.28354117	0.565838597	0.94849219
25	0	0.022117155	0.137818077	0.15993523
26	0.60935833	0.679453196	0.938907247	2.22771877
27	0.991139225	0.887094423	0.856831051	2.7350647
28	0.454320988	0.42694892	0.647714857	1.52898476
29	0.7284375	0.672993256	0.880438368	2.28186912
30	0.441899209	0.274093923	0.526656244	1.24264938
AVG	0.508240884	0.419176924	0.587407053	1.51482486
Median	0.623696688	0.453699467	0.614854191	1.85103809
MIN	0	0	0	0
MAX	0.991139225	0.887094423	0.970473258	2.7350647

Το **bpref** παίρνει τιμές κοντά στο 1 όταν τα σχετικά αρχεία επιστρέφονται πριν από τα μη σχετικά. Το **Avep** βγάζει τον μέσο όρο των precisions για κάθε θέση σχετικού αρχείου. Το DCG λαμβάνει υπόψη την σειρά που πήραμε τα έγγραφα και βγάζει τιμές κοντά στο 1 όταν τα σχετικά έγγραφα εμφανίζονται νωρίτερα από τα μη σχετικά, το **nDCG** βγάζει την αναλογία του dcg με το ιδανικό dcg (idcg).

Για τις μετρήσεις χρησιμοποίησαμε τους εξής τύπους:

Bpref:

$$bpref = \frac{1}{R} \sum_r 1 - \frac{|n \text{ ranked higher than } r|}{R}$$

Για R = relative documents retrieved

Avep:

$$Avep = \frac{1}{R} \sum_{1 \leq r \leq L} isrel(r)P(r)$$

Για isrel(r)= 0 ή 1

nDCG:

$$nDCG_p = \frac{DCG_p}{IDCG_p} \quad \mu\epsilon \quad DCG_p = \sum_{i=2}^p \frac{rel_i}{\log_2(i+1)}$$

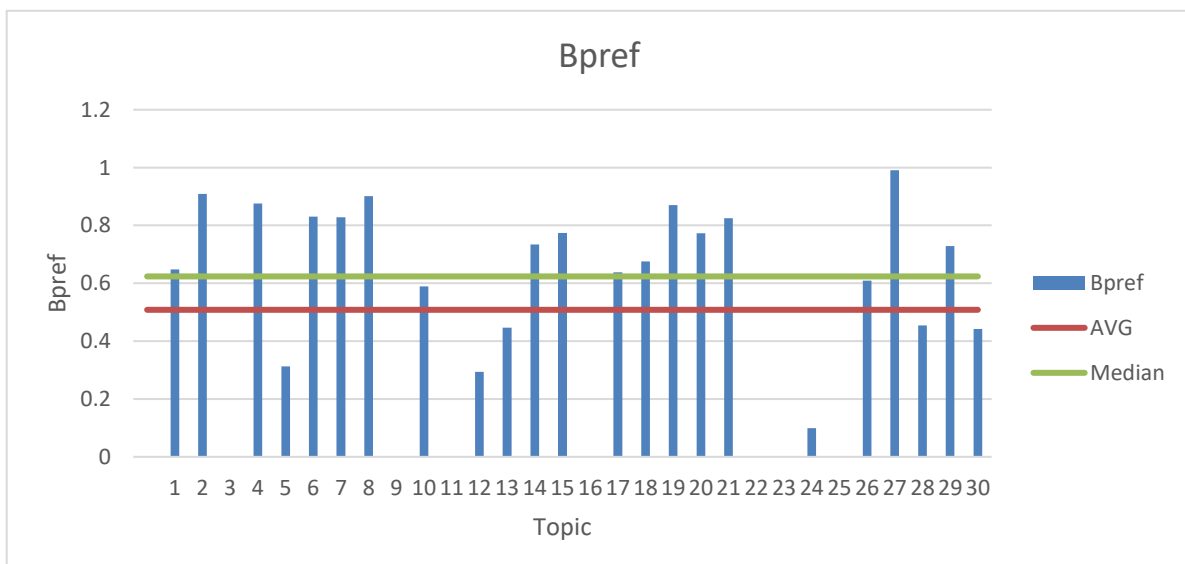
Για rel_i= τα σκορ των εγγράφων στο qrels.txt

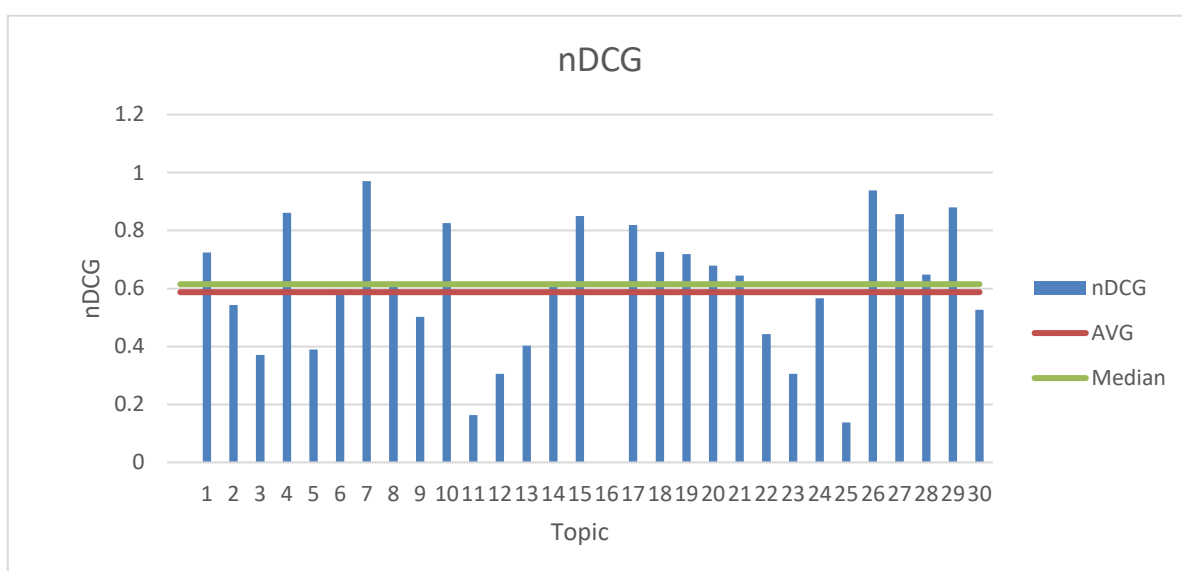
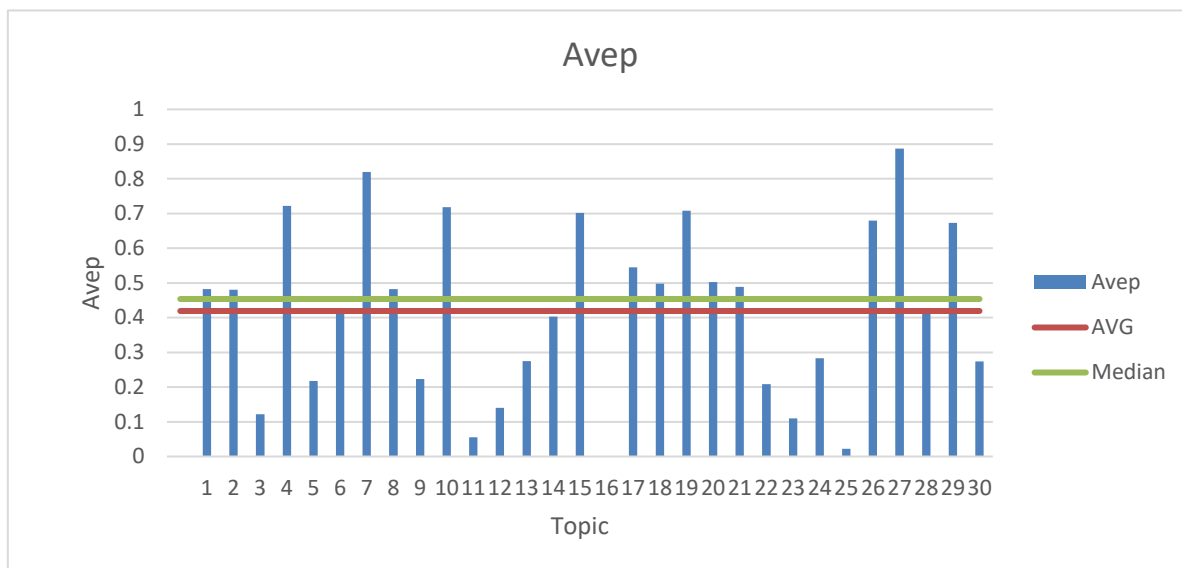
Παρατηρούμε ότι το σύστημα βγάζει καλές μετρήσεις στα topics 2, 4, 7, 8, 26, 27, 29. Το κοινό αυτών των queries ήταν ότι χρησιμοποιούσαν πολλούς και διαφορετικούς ιατρικούς όρους πράγμα που βοηθάει στο να βρεθούν περισσότερα σχετικά έγγραφα.

Εν αντιθέσει, τα topics 3, 9, 11, 16, 22, 23, 25 έχουν πολύ χαμηλές τιμές στις μετρήσεις. Παρατηρούμε ότι τα queries που παράγονται περιέχουν περισσότερες απλουστευμένες λέξεις παρά ιατρική ορολογία που αποτρέπει την εύρεση συγκεκριμένων συναφών εγγράφων λόγω της πληθώρας των εγγράφων που επιστρέφονται.

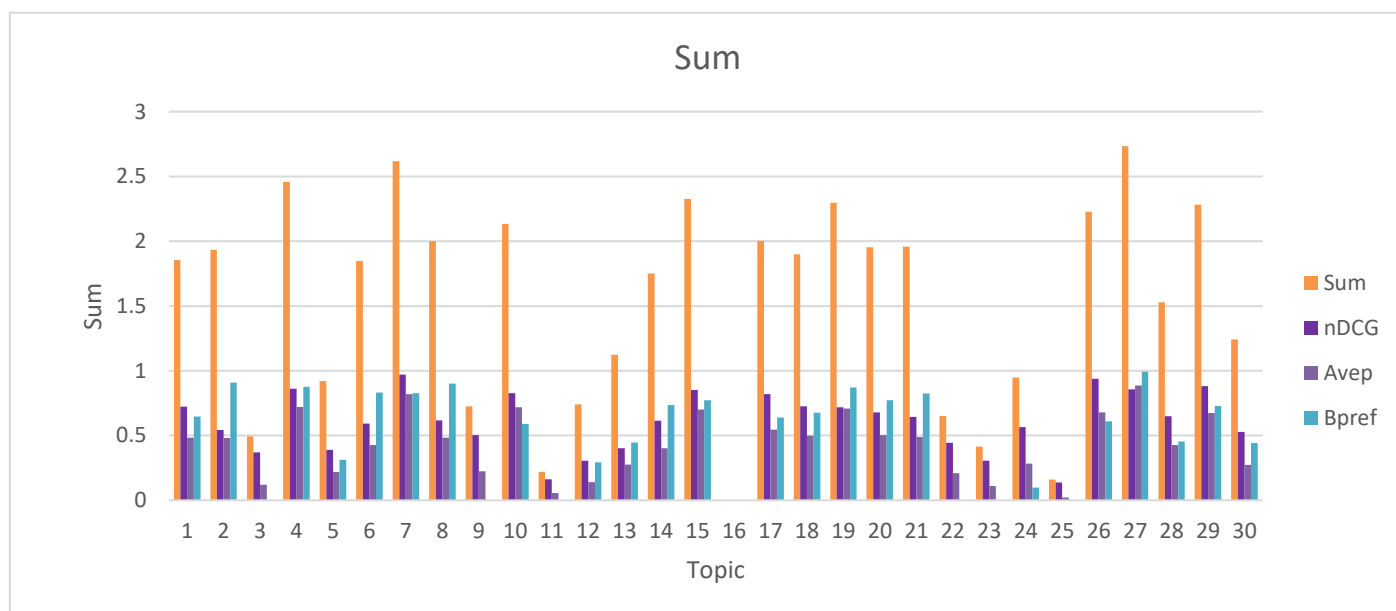
Σημείωση: Οι αρνητικές τιμές του bpref δεν προσμετρήθηκαν και απλοποιήθηκαν σε 0

Για τα στατιστικά:





Προσθέτοντας τις τρεις μετρικές βγάζουμε τα εξής αποτελέσματα:



Παρατηρούμε ότι οι τιμές των 3 μετρικών δεν διαφέρουν σημαντικά ανά topic.

4 Επίλογος

Τα αποτελέσματα ήταν αυτά που περιμέναμε κατά κύριο λόγο, ενώ ο υπολογισμός των μετρικών δίνει πολύ ενδιαφέροντα στατιστικά για την απόδοση ενός συστήματος.

Ωστόσο θα θέλαμε να μπορούσαμε να αφιερώσουμε περισσότερο χρόνο στο μάθημα για να μπορούσαμε να ολοκληρώσουμε το Interface και να κάνουμε πιο γρηγορά κάποιες διαδικασίες.

5 Αναφορές

Για των υπολογισμό των παραπάνω χρησιμοποιήθηκαν εκτός από το υλικό του μαθήματος και οι παρακάτω πηγές:

1. https://books.google.gr/books?id=aSEHCAAQBAJ&pg=PA251&lpg=PA251&dq=AveP+evaluation+metric&source=bl&ots=RZfkl-ENeB&sig=ACfU3U3zL_zyZbBW6qZ8LtTAQ6nR6eHd1A&hl=en&sa=X&ved=2ahUKEwiKwZH_nKXiAhXNxYUKHVa_Am0Q6AEwDHoECAgQAQ#v=onepage&q=AveP%20evaluation%20metric&f=false
2. <https://www.aclweb.org/anthology/M9341007>
3. <https://www.aclweb.org/anthology/M91-1002>
4. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.111.9845&rep=rep1&type=pdf>
5. <https://pdfs.semanticscholar.org/ae7f/19924ecb0f5b83a46117500d26b7675620df.pdf>
6. <https://www.aclweb.org/anthology/L16-1583>
7. https://www.jstage.jst.go.jp/article/ipsjdc/3/0/3_0_625/_pdf
8. [https://en.wikipedia.org/wiki/Evaluation_measures_\(information_retrieval\)](https://en.wikipedia.org/wiki/Evaluation_measures_(information_retrieval))
9. <https://trec.nist.gov/pubs/trec16/appendices/measures.pdf>
10. https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_489
11. <http://people.cs.georgetown.edu/~nazli/classes/ir-Slides/Evaluation-12.pdf>
12. <http://adcs-conference.org/2007/papers/21N.PDF>
13. <https://link.springer.com/article/10.1007/s10791-008-9059-7>