

αξιολόγηση με το εργαλείο trec_eval

Τώνια Κυριακοπούλου

Συστήματα Ανάκτησης Πληροφοριών
Ακαδημαϊκό Έτος 2020-2021

trec_eval

- Καθιερωμένο εργαλείο που χρησιμοποιείται από την κοινότητα του TREC για την αξιολόγηση της ανάκτησης.
 - Αρχείο με σωστές απαντήσεις (σχετικά κείμενα)
 - Αρχείο με ανακτηθέντα κείμενα – αποτελέσματα ανάκτησης
- Δημιουργήθηκε από τον Chris Buckley
- Υλοποίηση σε C
- Download: http://trec.nist.gov/trec_eval/index.html

Text REtrieval Conference (TREC)

*...to encourage research in information retrieval
from large text collections.*

NIST

National Institute of Standards and Technology

Εγκατάσταση

- C compiler
- Cygwin
- "make" στο source directory
- Οδηγίες:
 - https://github.com/usnistgov/trec_eval/blob/master/README.windows.md
- Έλεγχος:
 - Δείγμα αρχείων εισόδου και εξόδου περιλαμβάνεται στον κατάλογο test
 - "make quicktest"
 - Παραδείγματα από απλές αξιολογήσεις και συγκρίσεις αποτελεσμάτων
- **Σημαντικό: χρησιμοποιείτε το έτοιμο εκτελέσιμο αρχείο `trec_eval.exe` που δίνεται στο `eclass`!**

Σύνταξη εντολής του trec_eval

- Βασική:
 - `trec_eval` [-q] [-m measure] `rel_info_file` `your_results_file`
- -q flag
 - `trec_eval -q rel_info_file your_results_file`
 - Αποτελέσματα για κάθε ερώτημα ξεχωριστά
 - Περίληψη - Μέσος όρος από όλα τα ερωτήματα
- -m flag
 - `trec_eval -m official rel_info_file your_results_file`
 - Περίληψη - Μέσος όρος από όλα τα ερωτήματα (πχ. -m official, -m all_trec)
 - Περίληψη - Μέσος όρος από όλα τα ερωτήματα για μία μετρική (πχ. -m map)
- `rel_info_file` → `qrels`, οι σωστές απαντήσεις
- `your_results_file` → οι απαντήσεις του συστήματος ανάκτησής σου

Σωστές απαντήσεις – qrels

- Μορφή: (query_id iteration docno relevance)
- Παράδειγμα αρχείου `rel_info_file (qrels)`

query_id	iteration	docno	relevance
1	0	13	1
1	0	14	1
1	0	15	0
1	0	72	1
...			
2	0	80	0
2	0	90	1
...			

query_id=string
docno=string

Αποτελέσματα συστήματος ανάκτησης

- Μορφή: (query_id iter docno rank sim run_id)
- Παράδειγμα αρχείου [your_results_file](#)

q_id	iter	docno	rank	sim	run_id
01	0	2319	0	0.4035559	myIRmethod
01	0	2541	0	0.23646459	myIRmethod
01	0	1410	0	0.22437239	myIRmethod
...					
02	0	1591	0	0.21782021	myIRmethod
02	0	1605	0	0.20721076	myIRmethod
02	0	1827	0	0.2009484	myIRmethod

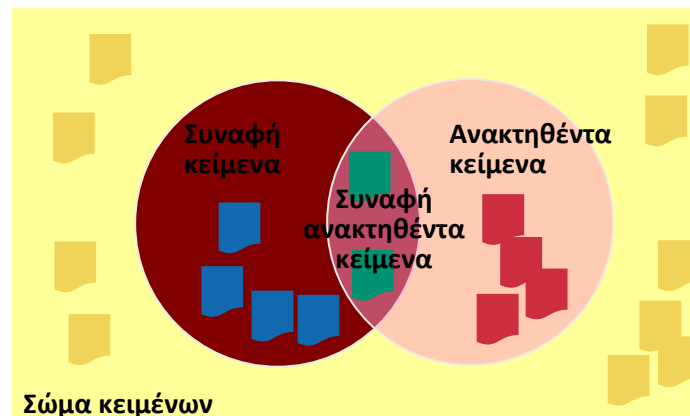
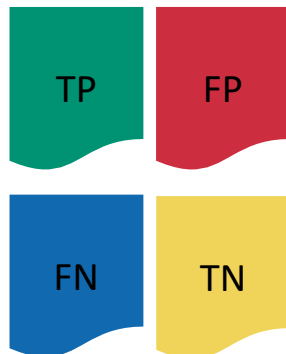
query_id=string
docno=string
sim=float
run_id=string (όνομα συστήματος ανάκτησης)
iter, rank: τα αγνοεί, απλά πρέπει να υπάρχουν

Έλεγχξε ότι...

- Τα 2 αρχεία είναι ταξινομημένα κατά query_id
- Το αρχείο αποτελεσμάτων είναι ταξινομημένο έτσι ώστε τα αποτελέσματα με το μεγαλύτερο similarity να δίνονται πρώτα (για κάθε query)
- Το trec_eval.exe και τα 2 αρχεία είναι στον ίδιο κατάλογο
- Διάβασες το README αρχείο!

Πώς γίνεται η αξιολόγηση...

- Εκτέλεσε από command line:
 - `trec_eval rel_info_file your_results_file [> eval.txt]`
- Διαβάζει όλα τα κείμενα που ανακτήθηκαν από το σύστημα ανάκτησης για ένα συγκεκριμένο ερώτημα από το `your_results_file` → `num_retrieved`
- Διαβάζει όλα τα σωστά σχετικά κείμενα για το ίδιο ερώτημα από το `rel_info_file (qrels)` → `num_relevant`
- Αναζητά τα ανακτηθέντα κείμενα μέσα στα σωστά για να δει πόσα από αυτά είναι σχετικά (relevant) → `num_relevant retrieved`



Μετρικές του trec_eval

runid	Name of the run (is the name given on the last field of the results file)
num_q	Total number of evaluated queries
num_ret	Total number of retrieved documents
num_rel	Total number of relevant documents (according to the qrels file)
num_rel_ret	Total number of relevant documents retrieved (in the results file)
map	Mean average precision (map)
gm_map	Average precision. Geometric mean
Rprec	Precision of the first R documents, where R are the number os relevants
bpref	Binary preference
recip_rank	Reciprical Rank
iprec_at_recall_0.00	Interpolated Recall - Precision Averages at 0.00 recall
iprec_at_recall_0.10	Interpolated Recall - Precision Averages at 0.10 recall
iprec_at_recall_0.20	Interpolated Recall - Precision Averages at 0.20 recall
iprec_at_recall_0.30	Interpolated Recall - Precision Averages at 0.30 recall
iprec_at_recall_0.40	Interpolated Recall - Precision Averages at 0.40 recall
iprec_at_recall_0.50	Interpolated Recall - Precision Averages at 0.50 recall
iprec_at_recall_0.60	Interpolated Recall - Precision Averages at 0.60 recall
iprec_at_recall_0.70	Interpolated Recall - Precision Averages at 0.70 recall
iprec_at_recall_0.80	Interpolated Recall - Precision Averages at 0.80 recall
iprec_at_recall_0.90	Interpolated Recall - Precision Averages at 0.90 recall
iprec_at_recall_1.00	Interpolated Recall - Precision Averages at 1.00 recall
P_5	Precision of the 5 first documents
P_10	Precision of the 10 first documents
P_15	Precision of the 15 first documents
P_20	Precision of the 20 first documents
P_30	Precision of the 30 first documents
P_100	Precision of the 100 first documents
P_200	Precision of the 200 first documents
P_500	Precision of the 500 first documents
P_1000	Precision of the 1000 first documents

Μετρικές του trec_eval

runid	all	STANDARD	
num_q	all	3	
num_ret	all	1500	
num_rel	all	561	
num_rel_ret	all	131	
map	all	0.1785	
gm_map	all	0.1051	
Rprec	all	0.2174	
bpref	all	0.1981	
recip_rank	all	0.4064	
iprec_at_recall_0.00	all		0.4665
iprec_at_recall_0.10	all		0.3884
iprec_at_recall_0.20	all		0.3186
iprec_at_recall_0.30	all		0.2852
iprec_at_recall_0.40	all		0.2666
iprec_at_recall_0.50	all		0.2184
iprec_at_recall_0.60	all		0.0822
iprec_at_recall_0.70	all		0.0348
iprec_at_recall_0.80	all		0.0312
iprec_at_recall_0.90	all		0.0312
iprec_at_recall_1.00	all		0.0312
P_5	all	0.2667	
P_10	all	0.3000	
P_15	all	0.3111	
P_20	all	0.3667	
P_30	all	0.3333	
P_100	all	0.2467	
P_200	all	0.1600	
P_500	all	0.0873	
P_1000	all	0.0437	