

بسم الله الرحمن الرحيم



دانشکده مهندسی برق و کامپیوتر

بازیابی هوشمند اطلاعات - تمرین اول

سید مهدی رضوی

استاد : خانم دکتر شاکری

مهر ماه ۱۴۰۲

فهرست مطالب

۲	۱	تمرین اول
۹	۲	تمرین دوم
۱۱	۳	تمرین سوم
۱۲	۴	تمرین چهارم
۱۲	۱.۴	R1
۱۲	۲.۴	R2
۱۳	۳.۴	R3

۱ تمرین اول

برای این تمرین ابتدا مقادیر متغیرهای b و k را با هم در پوشه‌های جداگانه مقایسه کردم و سپس به ارزیابی بهترین‌ها پرداختم.
برای مقایسه بین پارامترهای مختلف b و k اسکریپت‌هایی نوشتم که نتایج حالات مختلف را در فایل‌های متناظر می‌نویسیم و سپس با معیارهای ذکر شده به ارزیابی می‌پردازیم.

210	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
211	0.008	0.058	0.000	0.008	0.058	0.000	0.008	0.058	0.000	0.008	0.058
212	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
213	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
214	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
215	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
217	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
218	0.006	0.069	0.000	0.006	0.069	0.000	0.006	0.069	0.000	0.006	0.069
219	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
223	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
224	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
225	0.001	0.026	0.000	0.001	0.026	0.000	0.001	0.026	0.000	0.001	0.026
run-id	map	ndcg	p10								
/home/mahdi/Documents/bazyabi/runs/bm25-version-1/run1.txt	0.003	0.016	0.005								
/home/mahdi/Documents/bazyabi/runs/bm25-version-1/run2.txt	0.003	0.016	0.005								
/home/mahdi/Documents/bazyabi/runs/bm25-version-1/run3.txt	0.003	0.016	0.005								
/home/mahdi/Documents/bazyabi/runs/bm25-version-1/run4.txt	0.003	0.016	0.005								
Sig-Test: randomized, threshold set to 0.050000											
mahdi@MahdiRazavi:~/Documents/bazyabi/runs/bm25-version-1\$											

شکل ۱: تابع امتیازدهی اول

با توجه به این که تابع اول مستقل از پارامترهای ذکر شده هستند، در نتیجه نتایج کاملاً مشابه هم خواهد بود و نتایج به میزان منحصربفرد بودن یا همان IDF عبارات مشترک متن و پرس‌وجو خواهد بود.

File	Edit	View	Search	Terminal	Help
0.000	0.000	0.000	0.000		
211	0.002	0.035	0.000	0.001	0.033
0.000	0.000	0.000	0.000	0.000	0.000
212	0.016	0.135	0.000	0.016	0.150
0.000	0.010	0.119	0.000	0.035	0.143
213	0.016	0.128	0.000	0.035	0.143
0.100	0.033	0.137	0.100	0.008	0.077
214	0.015	0.094	0.000	0.008	0.077
0.000	0.000	0.000	0.000	0.000	0.000
215	0.007	0.099	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
217	0.001	0.029	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
218	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
219	0.002	0.049	0.000	0.001	0.026
0.000	0.001	0.028	0.000	0.001	0.026
223	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
224	0.003	0.049	0.000	0.014	0.076
0.100	0.063	0.160	0.100	0.100	0.009
225	0.010	0.112	0.000	0.002	0.045
0.000	0.005	0.069	0.000	0.000	0.000
run-id	map	ndcg	p10		
/home/mahdi/Documents/bazyabi/runs/bm25-version-2/run1.txt	0.009	0.052	0.017		
/home/mahdi/Documents/bazyabi/runs/bm25-version-2/run2.txt	0.009	0.052	0.013		
/home/mahdi/Documents/bazyabi/runs/bm25-version-2/run3.txt	0.009	0.054	0.016		
/home/mahdi/Documents/bazyabi/runs/bm25-version-2/run4.txt	0.008	0.051	0.015		
/home/mahdi/Documents/bazyabi/runs/bm25-version-2/run5.txt	0.007	0.045	0.012		
Sig-Test: randomized, threshold set to 0.050000					
mahdi@MahdiRazavi:~/Documents/bazyabi/runs/bm25-version-2\$					

شکل ۲: تابع امتیازدهی دوم

با توجه به ضابطه تابع امتیازدهی به اسناد ، باز متوجه خواهیم شد که ضابطه مستقل از پارامتر b می باشد. از تغییرات متغیر k متوجه خواهیم شد که هر چه میزان این متغیر به ۱ نزدیک تر باشد ، نتایج جستجو بهتر خواهد بود. زیرا که تاثیر فراوانی عبارات مشترک را کمتر خواهد کرد .

File	Edit	View	Search	Terminal	Help
0.029	0.000				
203	0.000	0.000	0.000	0.000	0.000
0.000	0.000				
204	0.000	0.000	0.000	0.000	0.000
0.000	0.000				
209	0.005	0.069	0.000	0.005	0.069
0.069	0.000				
210	0.000	0.000	0.000	0.000	0.000
0.000	0.000				
211	0.000	0.000	0.000	0.000	0.000
0.000	0.000				
212	0.036	0.222	0.000	0.036	0.222
0.222	0.000				
213	0.003	0.043	0.000	0.003	0.043
0.043	0.000				
214	0.000	0.000	0.000	0.000	0.000
0.000	0.000				
215	0.000	0.000	0.000	0.000	0.000
0.000	0.000				
217	0.000	0.000	0.000	0.000	0.000
0.000	0.000				
218	0.000	0.000	0.000	0.000	0.000
0.000	0.000				
219	0.000	0.000	0.000	0.000	0.000
0.000	0.000				
223	0.000	0.000	0.000	0.000	0.000
0.000	0.000				
224	0.010	0.066	0.000	0.010	0.066
0.066	0.000				
225	0.001	0.024	0.000	0.001	0.024
0.024	0.000				
run-id	map	ndcg	p10		
/home/mahdi/Documents/bazyabi/runs/bm25-version-3/run1.txt	0.004	0.026	0.006		
/home/mahdi/Documents/bazyabi/runs/bm25-version-3/run2.txt	0.004	0.026	0.006		
/home/mahdi/Documents/bazyabi/runs/bm25-version-3/run3.txt	0.004	0.026	0.006		
/home/mahdi/Documents/bazyabi/runs/bm25-version-3/run4.txt	0.004	0.026	0.006		
/home/mahdi/Documents/bazyabi/runs/bm25-version-3/run5.txt	0.004	0.026	0.006		
Sig-Test: randomized, threshold set to 0.050000					
mahdi@MahdiRazavi:~/Documents/bazyabi/runs/bm25-version-3\$					

شکل ۳: نتایج مرتب‌سازی با تابع امتیازدهی ۳

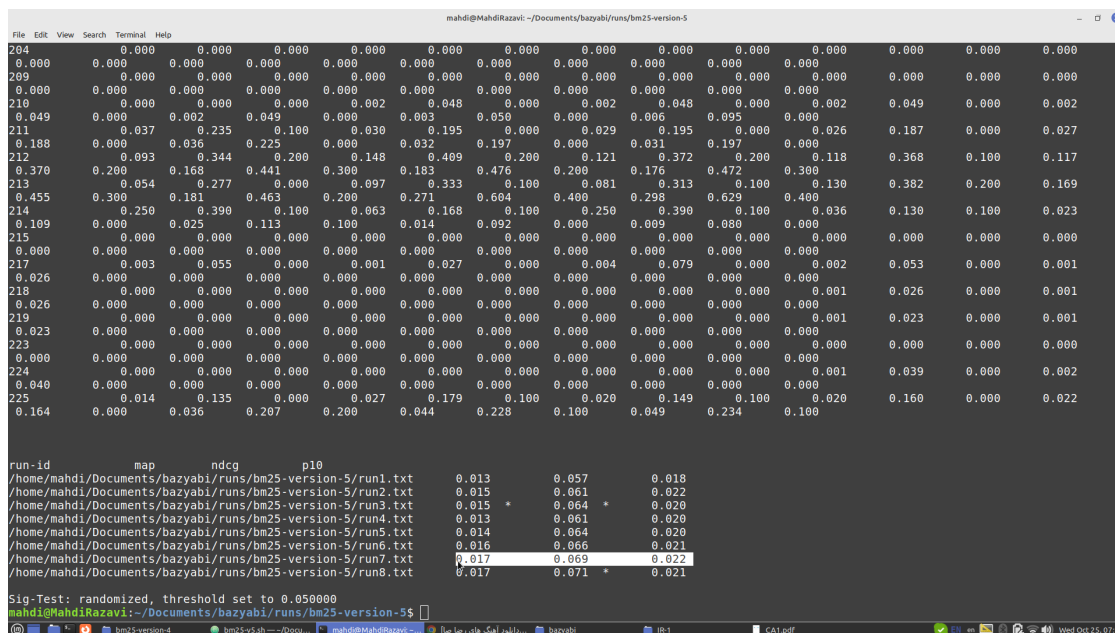
این روش نیز مانند روش اول مستقل از پارامترهای b و k خواهد بود. نتایج همگی مشابه هم خواهند بود. این روش در تلاش است که تاثیر طول اسناد ، تاثیر کمتری در نتایج نهایی داشته باشد. چرا که طبیعتا هر چه طول سند بیشتر باشد ، احتمال مرتبط بودن پرس‌وجو با آن بیشتر خواهد بود.

File	Edit	View	Search	Terminal	Help
203	0.000	0.001	0.000	0.030	0.000
0.031	0.000	0.001	0.000	0.027	0.000
204	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
209	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
210	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
211	0.000	0.049	0.260	0.100	0.130
0.245	0.100	0.127	0.393	0.100	0.420
212	0.062	0.289	0.200	0.097	0.335
0.312	0.100	0.122	0.370	0.200	0.355
213	0.050	0.263	0.000	0.104	0.355
0.394	0.200	0.147	0.400	0.300	0.272
214	0.262	0.462	0.100	0.098	0.272
0.241	0.100	0.056	0.215	0.100	0.264
215	0.000	0.000	0.000	0.000	0.466
0.000	0.000	0.000	0.000	0.000	0.100
217	0.002	0.053	0.000	0.002	0.138
0.055	0.000	0.000	0.000	0.000	0.320
218	0.000	0.000	0.000	0.000	0.100
0.000	0.000	0.000	0.000	0.000	0.054
219	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
223	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
224	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
225	0.000	0.102	0.000	0.022	0.174
0.161	0.000	0.025	0.171	0.000	0.016
run-id	map	ndcg	p10		
/home/mahdi/Documents/bazyabi/runs/bm25-version-4/run1.txt	0.013	0.060	0.018		
/home/mahdi/Documents/bazyabi/runs/bm25-version-4/run2.txt	0.015	0.070	0.018		
/home/mahdi/Documents/bazyabi/runs/bm25-version-4/run3.txt	0.016	0.068	0.018		
/home/mahdi/Documents/bazyabi/runs/bm25-version-4/run4.txt	0.015	0.066	0.019		
/home/mahdi/Documents/bazyabi/runs/bm25-version-4/run5.txt	0.014	0.064	0.021		
/home/mahdi/Documents/bazyabi/runs/bm25-version-4/run6.txt	0.015	0.066	0.021		
Sig-Test: randomized, threshold set to 0.050000					
mahdi@MahdiRazavi:~/Documents/bazyabi/runs/bm25-version-4\$					

شکل ۴: نتایج مرتب‌سازی با تابع امتیازدهی ۴

همانطور که در تصویر مشخص است ، بهترین نتایج مربوط به اجرای شماره ۳ می‌باشد. در این اجرا ما پارامترهای زیر را داشته‌ایم :

$$k = 1.75, b = 0.52$$



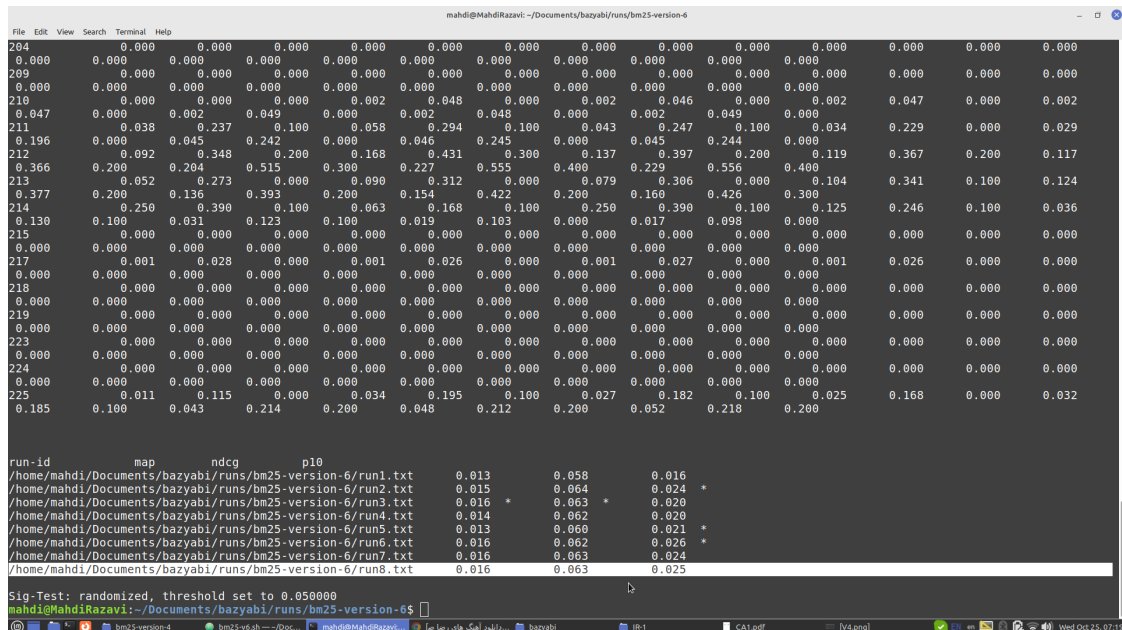
run-id	map	ndcg	p10
/home/mahdi/Documents/bazyabi/runs/bm25-version-5/run1.txt	0.013	0.057	0.018
/home/mahdi/Documents/bazyabi/runs/bm25-version-5/run2.txt	0.015	0.061	0.022
/home/mahdi/Documents/bazyabi/runs/bm25-version-5/run3.txt	0.015 *	0.064 *	0.020
/home/mahdi/Documents/bazyabi/runs/bm25-version-5/run4.txt	0.013	0.061	0.020
/home/mahdi/Documents/bazyabi/runs/bm25-version-5/run5.txt	0.014	0.064	0.020
/home/mahdi/Documents/bazyabi/runs/bm25-version-5/run6.txt	0.016	0.066	0.021
/home/mahdi/Documents/bazyabi/runs/bm25-version-5/run7.txt	0.017	0.069	0.022
/home/mahdi/Documents/bazyabi/runs/bm25-version-5/run8.txt	0.017	0.071 *	0.021

شکل ۵: نتایج مرتب‌سازی با تابع امتیازدهی ۵

در این روش طبق تصاویر نتایج مشاهده‌شده ، بهترین اجرای ما ، اجرای شماره ۷ خواهد بود. پارامترهای این اجرا عبارت خواهند بود از :

$$b = 0.92, k = 18.75$$

از این آزمایش نتیجه می‌گیریم که با افزایش مقدار k احتمال موفقیت‌آمیزتر بودن آزمایش ما بیشتر است. زیرا با افزایش این مقدار ، تاثیر عبارات بسیار تکرارشونده خیلی کم می‌شود.



شکل ۶: نتایج مرتب‌سازی با تابع امتیازدهی ۶

$k = 8.75, b = 0.82$

mahdi@MahdiRazavi: ~/Documents/bazyabi/runs/bm25-version-7																
File	Edit	View	Search	Terminal	Help											
204	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
209	0.005	0.069	0.069	0.000	0.005	0.069	0.000	0.000	0.005	0.069	0.000	0.005	0.069	0.000	0.005	0.005
0.069	0.000	0.005	0.069	0.000	0.005	0.069	0.000	0.000	0.005	0.069	0.000	0.005	0.069	0.000	0.005	0.005
210	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
211	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
212	0.036	0.222	0.000	0.036	0.222	0.000	0.036	0.222	0.000	0.036	0.222	0.000	0.036	0.222	0.000	0.036
0.222	0.000	0.036	0.222	0.000	0.036	0.222	0.000	0.036	0.222	0.000	0.036	0.222	0.000	0.036	0.222	0.000
213	0.003	0.043	0.000	0.003	0.043	0.000	0.003	0.043	0.000	0.003	0.043	0.000	0.003	0.043	0.000	0.003
0.043	0.000	0.003	0.043	0.000	0.003	0.043	0.000	0.003	0.043	0.000	0.003	0.043	0.000	0.003	0.000	0.000
214	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
215	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
217	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
218	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
219	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
223	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
224	0.010	0.066	0.000	0.010	0.066	0.000	0.010	0.066	0.000	0.010	0.066	0.000	0.010	0.066	0.000	0.010
0.066	0.000	0.010	0.066	0.000	0.010	0.066	0.000	0.010	0.066	0.000	0.010	0.066	0.000	0.010	0.066	0.000
225	0.001	0.024	0.000	0.001	0.024	0.000	0.001	0.024	0.000	0.001	0.024	0.000	0.001	0.024	0.000	0.001
0.024	0.000	0.001	0.024	0.000	0.001	0.024	0.000	0.001	0.024	0.000	0.001	0.024	0.000	0.001	0.024	0.000
run-id map ndcg p10																
/home/mahdi/Documents/bazyabi/runs/bm25-version-7/run1.txt	0.004	0.026	0.006													
/home/mahdi/Documents/bazyabi/runs/bm25-version-7/run2.txt	0.004	0.026	0.006													
/home/mahdi/Documents/bazyabi/runs/bm25-version-7/run3.txt	0.004	0.026	0.006													
/home/mahdi/Documents/bazyabi/runs/bm25-version-7/run4.txt	0.004	0.026	0.006													
/home/mahdi/Documents/bazyabi/runs/bm25-version-7/run5.txt	0.004	0.026	0.006													
/home/mahdi/Documents/bazyabi/runs/bm25-version-7/run6.txt	0.004	0.026	0.006													
/home/mahdi/Documents/bazyabi/runs/bm25-version-7/run7.txt	0.004	0.026	0.006													
/home/mahdi/Documents/bazyabi/runs/bm25-version-7/run8.txt	0.004	0.026	0.006													
Sig-Test: randomized, threshold set to 0.050000																
mahdi@MahdiRazavi:~/Documents/bazyabi/runs/bm25-version-7\$																

شکل ۷: نتایج مرتب‌سازی با تابع امتیازدهی ۷

متاسفانه تغییری در نتایج آزمایش‌ها مشاهده نشد. چون که آزمایش‌ها مستقل از k می‌باشند، متوجه خواهیم شد که تاثیر پارامتر b نیز بسیار ناچیز است.

۲ تمرین دوم

همانند سوال ۱ اسکرپت‌های توابع امتیازدهی به اسناد را می‌نویسیم.

```

mahdi@MahdiRazavi:~/Documents/bazyabi/runs/pln_1
File Edit View Search Terminal Help
Caused by: java.lang.reflect.InvocationTargetException
    at jdk.internal.reflect.GeneratedConstructorAccessor1.newInstance(Unknown Source)
    at java.base/jdk.internal.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
    at java.base/java.lang.reflect.Constructor.newInstance(Constructor.java:490)
    at org.lemurproject.galago.core.retrieval.FeatureFactory.getIterator(FeatureFactory.java:381)
    ... 9 more
Caused by: java.lang.ArithmeticException: / by zero
    at org.lemurproject.galago.core.retrieval.iterator.scoring.PLN_1.score(PLN_1.java:69)
    at org.lemurproject.galago.core.retrieval.iterator.scoring.PLN_1.<init>(PLN_1.java:59)
    ... 13 more
gids      /home/mahdi/Documents/bazyabi/runs/pln_1/run1.txt-map /home/mahdi/Documents/bazyabi/runs/pln_1/run1.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_1/run1.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_1/run2.txt-map /home/mahdi/Documents/bazyabi/runs/pln_1/run2.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_1/run2.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_1/run3.txt-map /home/mahdi/Documents/bazyabi/runs/pln_1/run3.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_1/run3.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_1/run4.txt-map /home/mahdi/Documents/bazyabi/runs/pln_1/run4.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_1/run4.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_1/run5.txt-map /home/mahdi/Documents/bazyabi/runs/pln_1/run5.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_1/run5.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_1/run6.txt-map /home/mahdi/Documents/bazyabi/runs/pln_1/run6.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_1/run6.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_1/run7.txt-map /home/mahdi/Documents/bazyabi/runs/pln_1/run7.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_1/run7.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_1/run8.txt-map /home/mahdi/Documents/bazyabi/runs/pln_1/run8.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_1/run8.txt-p10
1          0.098      0.329      0.000      0.098      0.329      0.000      0.098      0.329      0.000      0.098      0.329      0.000      0.098      0.329      0.000      0.098
0.329      0.000      0.098      0.329      0.000      0.000      0.098      0.329      0.000      0.000      0.098      0.329      0.000      0.098      0.329      0.000      0.098
2          0.142      0.025      0.000      0.142      0.025      0.000      0.142      0.025      0.000      0.142      0.025      0.000      0.142      0.025      0.000      0.142
0.142      0.000      0.025      0.142      0.000      0.000      0.025      0.142      0.000      0.000      0.025      0.142      0.000      0.025      0.142      0.000      0.025
4          0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000
0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000
8          0.119      0.013      0.000      0.119      0.013      0.000      0.119      0.013      0.000      0.119      0.013      0.000      0.119      0.013      0.000      0.119      0.013
0.119      0.000      0.013      0.119      0.000      0.000      0.013      0.119      0.000      0.000      0.013      0.119      0.000      0.013      0.119      0.000      0.013
12         0.053      0.002      0.000      0.053      0.002      0.000      0.053      0.002      0.000      0.053      0.002      0.000      0.053      0.002      0.000      0.053      0.002
0.053      0.000      0.002      0.053      0.000      0.000      0.002      0.053      0.000      0.000      0.002      0.053      0.000      0.002      0.053      0.000      0.002

run-id      map      ndcg      p10
/home/mahdi/Documents/bazyabi/runs/pln_1/run1.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_1/run2.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_1/run3.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_1/run4.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_1/run5.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_1/run6.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_1/run7.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_1/run8.txt      0.028      0.129      0.000
Sig-Test: randomized, threshold set to 0.050000

```

شکل ۸: تاثیر تابع تبدیل

متاسفانه تغییری در نتایج آزمایش‌ها مشاهده نشد. چون که آزمایش‌ها مستقل از k می‌باشند، متوجه خواهیم شد که تاثیر پارامتر b نیز بسیار ناچیز است.

```
File Edit View Search Terminal Help
mahdi@MahdiRazavi: ~/Documents/bazyabi/runs/pln_2
at java.lang.reflect.GeneratedConstructorAccessor1.newInstance(Unknown Source)
at java.base/java.lang.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
at java.base/java.lang.reflect.Constructor.newInstance(Constructor.java:490)
at org.lemurproject.galago.core.retrieval.FeatureFactory.getIterator(FeatureFactory.java:381)
... 9 more
Caused by: java.lang.ArithmeticException: / by zero
at org.lemurproject.galago.core.retrieval.iterator.scoring.PLN_2.score(PLN_2.java:69)
at org.lemurproject.galago.core.retrieval.iterator.scoring.PLN_2.<init>(PLN_2.java:59)
... 13 more
qids      /home/mahdi/Documents/bazyabi/runs/pln_2/run1.txt-map /home/mahdi/Documents/bazyabi/runs/pln_2/run1.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_2/run1.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_2/run2.txt-map /home/mahdi/Documents/bazyabi/runs/pln_2/run2.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_2/run2.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_2/run3.txt-map /home/mahdi/Documents/bazyabi/runs/pln_2/run3.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_2/run3.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_2/run4.txt-map /home/mahdi/Documents/bazyabi/runs/pln_2/run4.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_2/run4.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_2/run5.txt-map /home/mahdi/Documents/bazyabi/runs/pln_2/run5.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_2/run5.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_2/run6.txt-map /home/mahdi/Documents/bazyabi/runs/pln_2/run6.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_2/run6.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_2/run7.txt-map /home/mahdi/Documents/bazyabi/runs/pln_2/run7.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_2/run7.txt-p10 /home/mahdi/Documents/bazyabi/runs/pln_2/run8.txt-map /home/mahdi/Documents/bazyabi/runs/pln_2/run8.txt-ndcg /home/mahdi/Documents/bazyabi/runs/pln_2/run8.txt-p10
1          0.098      0.329      0.000      0.098      0.329      0.000      0.098      0.329      0.000      0.098      0.329      0.000      0.098
0.329      0.000      0.098      0.329      0.000      0.098      0.329      0.000      0.098      0.329      0.000      0.098      0.329      0.000
2          0.025      0.142      0.000      0.025      0.142      0.000      0.025      0.142      0.000      0.025      0.142      0.000      0.025
0.142      0.000      0.025      0.142      0.000      0.025      0.142      0.000      0.025      0.142      0.000      0.025      0.142      0.000
4          0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000
0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000
8          0.013      0.119      0.000      0.013      0.119      0.000      0.013      0.119      0.000      0.013      0.119      0.000      0.013
0.119      0.000      0.013      0.119      0.000      0.013      0.119      0.000      0.013      0.119      0.000      0.013      0.119      0.000
12         0.002      0.053      0.000      0.002      0.053      0.000      0.002      0.053      0.000      0.002      0.053      0.000      0.002
0.053      0.000      0.002      0.053      0.000      0.002      0.053      0.000      0.002      0.053      0.000      0.002      0.053      0.000

run-id      map      ndcg      p10
/home/mahdi/Documents/bazyabi/runs/pln_2/run1.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_2/run2.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_2/run3.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_2/run4.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_2/run5.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_2/run6.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_2/run7.txt      0.028      0.129      0.000
/home/mahdi/Documents/bazyabi/runs/pln_2/run8.txt      0.028      0.129      0.000

Sig-Test: randomized, threshold set to 0.050000
mahdi@MahdiRazavi:~/Documents/bazyabi/runs/pln_2$
```

شکل ۹: تاثیر تابع تبدیل ۲

متاسفانه تغییری در نتایج آزمایش‌ها مشاهده نشد. چون که آزمایش‌ها مستقل از k می‌باشند، متوجه خواهیم شد که تاثیر پارامتر b نیز بسیار ناچیز است.

۳ تمرین سوم

$$\vec{q}_m = \alpha \vec{q} + \frac{\beta}{D_r} \sum_{\vec{d}_j \in D_r} \vec{d}_j - \frac{\gamma}{D_{nr}} \sum_{\vec{d}_j \in D_{nr}} \vec{d}_j$$

در گام نخست پارامترها را طبق خواسته سوال برابر یک در نظر خواهیم گرفت.

$$\alpha = 1, \beta = 1, \gamma = 1 \quad (1)$$

سپس به محاسبه کوئری اصلاح شده می‌پردازیم.

$$\vec{q}_0 = (0, 0, 0, 0, 1, 1, 1)$$

$$\vec{d}_{R1} = (0, 0, 0, 1, 0, 1, 2), \vec{d}_{R2} = (0, 0, 1, 0, 1, 1, 1)$$

$$\vec{d}_{NR1} = (1, 1, 1, 0, 1, 0, 0), \vec{d}_{NR2} = (1, 1, 0, 1, 1, 0, 0)$$

$$\vec{q}_1 = \alpha \vec{q}_0 + \frac{\beta}{D_r} \sum_{\vec{d}_j \in D_r} \vec{d}_j - \frac{\gamma}{D_{nr}} \sum_{\vec{d}_j \in D_{nr}} \vec{d}_j$$

$$\vec{q}_1 = (0, 0, 0, 0, 0.5, 2, 2.5)$$

در درجه‌ی اول بعد از محاسبه میزان شباهت کوئری اصلاح شده و داکيومنت‌ها ، متوجه خواهیم شد که اسناد relevant فاصله قابل توجه‌تری با اسناد non-relevant خواهند گرفت.

$$Sim(q_0, d_{R1}) = 3, Sim(q_0, d_{R2}) = 3, Sim(q_0, d_{NR1}) = 1, Sim(q_0, d_{NR2}) = 1$$

$$Sim(q_1, d_{R1}) = 7, Sim(q_1, d_{R2}) = 5, Sim(q_1, d_{NR1}) = 0.5, Sim(q_1, d_{NR2}) = 0.5$$

همچنین با توجه به بهبود میزان شباهت بین کوئری اول و دوم ، فاصله گرفتن اسناد مرتبط و نامرتبط ، به بهبود کوئری می‌پردازیم.

$$\vec{q}_2 = \alpha \vec{q}_1 + \frac{\beta}{D_r} \sum_{\vec{d}_j \in D_r} \vec{d}_j - \frac{\gamma}{D_{nr}} \sum_{\vec{d}_j \in D_{nr}} \vec{d}_j$$

$$\vec{q}_2 = (0, 0, 0.35, 0.35, 0.85, 2.8, 3.7)$$

$$Sim(q_2, d_{R1}) = 13.9, Sim(q_2, d_{R2}) = 7.7, Sim(q_2, d_{NR1}) = 1.2, Sim(q_2, d_{NR2}) = 1.2$$

۴ تمرین چهارم

R1 ۱.۴

با توجه به این که در این سیستم‌ها فقط یک پرس‌وجو داریم ، ($n = 1$) در نتیجه مقادیر MAP و GMAP با هم برابر خواهند بود.

$$Precision = \frac{3}{5}$$

$$Recall = \frac{3}{10}$$

$$P@4 = \frac{3}{5}$$

$$MAP = GMAP = Average - Precision_1 = 0.227$$

R2 ۲.۴

$$Precision = \frac{10}{20} = 0.5$$

$$Recall = \frac{10}{10} = 1$$

$$P@4 = \frac{2}{4} = 0.5, P@7 = \frac{3}{7}, P@12 = \frac{6}{12} = 0.5$$

$$R - Precision = \frac{5}{10} = 0.5$$

$$MAP = GMAP = \frac{1 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.54 + 0.53 + 0.56 + 0.5}{10} = 0.563$$

$$Precision = \frac{5}{11}$$

$$Recall = \frac{5}{10} = 0.5$$

$$P@4 = \frac{2}{4}, P@7 = \frac{3}{7}$$

$$R - Precision = \frac{5}{10}$$

$$MAP = GMAP = Average - Precision_1 = \frac{1 + 0.5 + 0.6 + 0.5 + 0.5}{10} = 0.31$$