

به نام خدا

گزارش تمرین برنامه نویسی اول

زهره دهقانپان-۹۸۱۳۱۰۵۹

بخش دوم)

در این بخش بردار TF-IDF را هم به صورت باینری و هم به صورت عددی محاسبه کردیم و برای معیار فاصله نیز هر دو معیار جاکارد و کسینوسی را استفاده کردیم. نتایج doc های مرتبط با هر کویری و خروجی این مرحله به صورت زیر است:

```
C:\Users\zdehg\AppData\Local\Programs\Python\Python37\python.exe
"D:/univesity/foqelisans/information retrieval/HW1_98131059/TF_IDF.py"
C:\Users\zdehg\AppData\Local\Programs\Python\Python37\lib\site-
packages\sklearn\externals\joblib\externals\cloudpickle\cloudpickle.py:47:
DeprecationWarning: the imp module is deprecated in favour of importlib; see the module's
documentation for alternative uses
import imp
parsing file finished
number of docs: 4600
number of distinct words: 72786
avg length of docs: 355.8517391304348
doc with max length: 1178s1
doc with min length: 3109s1
calculating IDF finished
0
calculating TF-IDF finished
*****numeric part*****
15 doc similar to query with cosine distance is :
[('41867s1', 0.40768069214406205), ('45702s7', 0.37465801623905215), ('5318s1',
0.37418807745449845), ('51230s8', 0.37171554660605083), ('41732s4',
0.34963294841508535), ('47137s3', 0.34500124495238266), ('42870s2',
0.34207082666386396), ('45792s2', 0.3101701475079292), ('433s1', 0.306020788666079),
('51890s4', 0.30470017092154134), ('38384s1', 0.3038827406667222), ('42190s4',
0.2928908784622224), ('2181s1', 0.2873772802214949), ('52260s2', 0.286573988319667),
('14672s1', 0.28635056975559303)]
15 doc similar to query 7 with cosine distance is :
[('41408s2', 0.45806054331972573), ('50542s5', 0.4291851223423028), ('55779s3',
0.42734691718011886), ('45596s1', 0.3979464765931685), ('48163s5',
0.38708040334517524), ('41115s3', 0.37262889957910394), ('45375s1',
0.3598297338446199), ('46314s2', 0.35622059886530905), ('52879s3', 0.3464836680940482),
('46602s6', 0.34236120127292735), ('58534s7', 0.3400348572734495), ('50291s4',
0.33919722816192993), ('50780s3', 0.33838742815703066), ('22483s1',
0.3363985418059458), ('53988s8', 0.32250475708257315)]
15 doc similar to query 8 with cosine distance is :
[('16816s1', 0.39148182004028864), ('40892s2', 0.3499791178667865), ('44195s1',
0.34903875900528564), ('44206s1', 0.33067126777992994), ('44182s1',
0.3299281681051315), ('54479s3', 0.3227221170810966), ('51034s1', 0.32122899459569615),
('47559s2', 0.314259864332172), ('16730s1', 0.3040007856351336), ('40706s1',
0.3001640857275746), ('54524s3', 0.29965126891026034), ('54292s1', 0.2983696808670272),
('54463s1', 0.29806624118523334), ('47650s1', 0.29735741445674707), ('47659s3',
0.29727007188583016)]
15 doc similar to query 9 with cosine distance is :
[('54407s1', 0.3725465601648833), ('4452s1', 0.37131752357487025), ('52036s1',
0.36678249446532474), ('54365s1', 0.3579199505378723), ('33454s1', 0.3571017624242026),
('126s1', 0.35083464228375577), ('3220s1', 0.34908654159372743), ('17370s1',
0.3460665696951601), ('2462s1', 0.34529843043766345), ('54463s1', 0.3430239594515186),
('54446s2', 0.3410744186654833), ('2269s1', 0.32992142387613027), ('42766s2',
0.3271853286967616), ('47363s1', 0.3227355141645786), ('767s1', 0.3222999726249234)]
15 doc similar to query 10 with cosine distance is :
```

```
[('6026s1', 0.5918835512310762), ('12326s1', 0.5881534494168745), ('54104s7', 0.5737501145804448), ('49336s2', 0.5701503388651202), ('35329s1', 0.5659939745310274), ('44518s6', 0.5599391717735447), ('45133s2', 0.5502180290160907), ('2833s1', 0.5487471878303067), ('53585s1', 0.5468608118673381), ('44931s1', 0.5425330617116667), ('43981s3', 0.5418619826198917), ('46817s4', 0.5228234899840969), ('45458s4', 0.5125739408300357), ('33972s1', 0.5093705702194234), ('43981s4', 0.5018573921577011)]
15 doc similar to query with jaccard distance is :
[('3347s1', 0.09523809523809523), ('2162s1', 0.07936507936507936), ('2484s1', 0.07142857142857142), ('2801s1', 0.07142857142857142), ('2940s1', 0.06521739130434782), ('48s1', 0.06382978723404255), ('3810s1', 0.0625), ('1492s1', 0.061224489795918366), ('3456s1', 0.061224489795918366), ('2433s1', 0.06060606060606061), ('1669s1', 0.060240963855421686), ('458s1', 0.06), ('2635s1', 0.05970149253731343), ('2405s1', 0.058823529411764705), ('2786s1', 0.058823529411764705)]
15 doc similar to query 7 with jaccard distance is :
[('3869s1', 0.09259259259259259), ('1164s1', 0.08928571428571429), ('597s1', 0.08), ('1010s1', 0.07692307692307693), ('2561s1', 0.07407407407407407), ('26377s1', 0.07142857142857142), ('1577s1', 0.06976744186046512), ('2603s1', 0.06896551724137931), ('2084s1', 0.06779661016949153), ('2359s1', 0.06779661016949153), ('3413s1', 0.06779661016949153), ('3992s1', 0.06666666666666667), ('4062s1', 0.06666666666666667), ('2525s1', 0.06557377049180328), ('200s1', 0.06382978723404255)]
15 doc similar to query 8 with jaccard distance is :
[('3831s1', 0.09230769230769231), ('2049s1', 0.08035714285714286), ('3437s1', 0.072), ('1597s1', 0.0660377358490566), ('3581s1', 0.06557377049180328), ('3425s1', 0.06542056074766354), ('541s1', 0.06481481481481481), ('33519s1', 0.06363636363636363), ('1882s1', 0.06349206349206349), ('499s1', 0.06060606060606061), ('1157s1', 0.06), ('2485s1', 0.05982905982905983), ('1371s1', 0.05952380952380952), ('42729s5', 0.056451612903225805), ('2532s1', 0.056074766355140186)]
15 doc similar to query 9 with jaccard distance is :
[('58534s7', 0.061946902654867256), ('3362s1', 0.06164383561643835), ('3870s1', 0.06164383561643835), ('2851s1', 0.059880239520958084), ('1920s1', 0.05759162303664921), ('1315s1', 0.05714285714285714), ('1764s1', 0.05442176870748299), ('3117s1', 0.05343511450381679), ('1762s1', 0.052980132450331126), ('668s1', 0.05263157894736842), ('1986s1', 0.05232558139534884), ('89s1', 0.05228758169934641), ('3260s1', 0.05188679245283019), ('299s1', 0.051470588235294115), ('2005s1', 0.051094890510948905)]
15 doc similar to query 10 with jaccard distance is :
[('138s1', 0.07526881720430108), ('2210s1', 0.0707070707070707), ('3033s1', 0.0684931506849315), ('4017s1', 0.0625), ('402s1', 0.06172839506172839), ('1158s1', 0.06060606060606061), ('1086s1', 0.057971014492753624), ('3710s1', 0.057971014492753624), ('3693s1', 0.053763440860215055), ('642s1', 0.05333333333333334), ('1167s1', 0.05263157894736842), ('46028s4', 0.05223880597014925), ('2755s1', 0.05194805194805195), ('3456s1', 0.05172413793103448), ('53012s3', 0.05128205128205128)]
*****binary part*****
15 doc similar to query with cosine distance is :
[('3313s1', 0.16220358710558688), ('2166s1', 0.1571766602286585), ('766s1', 0.13510750178457392), ('1344s1', 0.13415335337454767), ('38384s1', 0.1331442915106724), ('13687s1', 0.12954473791676183), ('21181s1', 0.12492164692702841), ('30496s1', 0.12117180412091758), ('13s1', 0.11884203638966252), ('1669s1', 0.11678646509992432), ('437s1', 0.11565548706921609), ('50838s8', 0.11540942475410401), ('596s1', 0.11402794900195579), ('42324s1', 0.1138564754270647), ('16629s1', 0.11081376051195968)]
15 doc similar to query 7 with cosine distance is :
[('53012s3', 0.1865850796380645), ('23990s1', 0.1820007200813376), ('48163s5', 0.1651444892394699), ('46314s2', 0.15635718973681165), ('53209s4', 0.15420073369069393), ('1430s1', 0.15408314005570548), ('52190s5', 0.1450540307552117), ('52879s3', 0.13711277804261876), ('43906s11', 0.136275717762096), ('50542s5', 0.1361903077349649), ('53408s7', 0.13610720081174268), ('51129s4', 0.1342294968842063), ('22483s1', 0.13172415667655657), ('22369s1', 0.13091207973015265), ('3041s1', 0.12833889650978872)]
15 doc similar to query 8 with cosine distance is :
[('50703s1', 0.1620995220657764), ('50542s5', 0.14633441405508568), ('1215s1', 0.13214279756592728), ('4017s1', 0.12313579067177151), ('1302s1', 0.1221692791312313),
```

```
('499s1', 0.12069305044828191), ('54212s5', 0.1204612953700747), ('1987s1',  
0.11854478615281262), ('50822s7', 0.11580852609083757), ('51029s5',  
0.11525669678757333), ('3873s1', 0.11498137627443339), ('1220s1', 0.11416884817969981),  
('992s1', 0.11380345003975015), ('1241s1', 0.11256998316988902), ('48560s4',  
0.11160946728515911)]  
15 doc similar to query 9 with cosine distance is :  
[('54446s2', 0.13322861605658842), ('58534s7', 0.129202056441069), ('3105s1',  
0.12763213487813382), ('3154s1', 0.12344039302250714), ('54407s1',  
0.12194169831368663), ('2398s1', 0.12187231528178193), ('2315s1', 0.11788421750615917),  
('47363s1', 0.11661059156832004), ('3624s1', 0.11584123631185077), ('3230s1',  
0.11550226283549897), ('1381s1', 0.11504710089035261), ('1764s1', 0.11366029494097096),  
('2803s1', 0.113136178503411), ('757s1', 0.1110994264098954), ('871s1',  
0.11089755906075416)]  
15 doc similar to query 10 with cosine distance is :  
[('21843s1', 0.17303363729317964), ('1288s1', 0.17077903475276612), ('4017s1',  
0.1659084860206198), ('12326s1', 0.1499352840710122), ('46028s4', 0.14853763260739017),  
('46205s2', 0.1436685433709042), ('53585s1', 0.14275989289199453), ('43030s1',  
0.14081495078465098), ('138s1', 0.13884002854635533), ('12707s1', 0.13660006312439382),  
('43981s4', 0.13552922066689616), ('6026s1', 0.1354042596231545), ('41921s1',  
0.1336694498325998), ('42242s3', 0.13356292125405644), ('42196s1',  
0.1322441867596201)]  
15 doc similar to query with jaccard distance is :  
[('2166s1', 0.13114754098360656), ('3347s1', 0.11904761904761904), ('2250s1',  
0.10526315789473684), ('1577s1', 0.10204081632653061), ('3456s1', 0.10204081632653061),  
('2339s1', 0.1), ('1487s1', 0.09803921568627451), ('1344s1', 0.09782608695652174),  
('21181s1', 0.0975609756097561), ('3191s1', 0.0967741935483871), ('30496s1',  
0.0967741935483871), ('1669s1', 0.0963855421686747), ('2162s1', 0.09523809523809523),  
('3810s1', 0.09375), ('2084s1', 0.09230769230769231)]  
15 doc similar to query 7 with jaccard distance is :  
[('53012s3', 0.12698412698412698), ('175s1', 0.12244897959183673), ('53209s4',  
0.11940298507462686), ('3271s1', 0.11904761904761904), ('1577s1', 0.11627906976744186),  
('2822s1', 0.11627906976744186), ('48163s5', 0.11290322580645161), ('43906s11',  
0.11290322580645161), ('50542s5', 0.11267605633802817), ('3869s1', 0.1111111111111111),  
('2827s1', 0.10909090909090909), ('499s1', 0.10810810810810811), ('46314s2',  
0.10714285714285714), ('23990s1', 0.10526315789473684), ('1993s1', 0.1038961038961039)]  
15 doc similar to query 8 with jaccard distance is :  
[('50542s5', 0.14285714285714285), ('3530s1', 0.125), ('3831s1', 0.12307692307692308),  
('2304s1', 0.12195121951219512), ('499s1', 0.12121212121212122), ('2827s1',  
0.11904761904761904), ('48560s4', 0.11904761904761904), ('851s1', 0.11827956989247312),  
('658s1', 0.11764705882352941), ('4017s1', 0.1171875), ('50822s7',  
0.11702127659574468), ('2049s1', 0.11607142857142858), ('815s1', 0.11458333333333333),  
('2484s1', 0.11392405063291139), ('2433s1', 0.11363636363636363)]  
15 doc similar to query 9 with jaccard distance is :  
[('58534s7', 0.12389380530973451), ('54446s2', 0.11731843575418995), ('2398s1',  
0.11029411764705882), ('1921s1', 0.10948905109489052), ('1764s1', 0.10884353741496598),  
('2876s1', 0.10810810810810811), ('2538s1', 0.10759493670886076), ('3230s1', 0.10625),  
('871s1', 0.1043956043956044), ('22369s1', 0.10179640718562874), ('2803s1',  
0.10071942446043165), ('2115s1', 0.1), ('2147s1', 0.1), ('2782s1', 0.1), ('658s1',  
0.09929078014184398)]  
15 doc similar to query 10 with jaccard distance is :  
[('54306s8', 0.125), ('21843s1', 0.11504424778761062), ('1288s1', 0.11363636363636363),  
('46205s2', 0.11363636363636363), ('53585s1', 0.11111111111111111), ('12707s1',  
0.10810810810810811), ('42242s3', 0.10784313725490197), ('138s1', 0.10752688172043011),  
('4017s1', 0.10714285714285714), ('50542s5', 0.10714285714285714), ('43981s4',  
0.10344827586206896), ('2451s1', 0.10144927536231885), ('42652s3',  
0.10091743119266056), ('6026s1', 0.1), ('49925s3', 0.1)]
```

Process finished with exit code 0

مقایسه معیارهای فاصله و روش محاسبه بردار TF-IDF به این صورت است که همان طور که در بخش چهارم می بینید، استفاده از فاصله کسینوسی در هر ۴ معیار، نتایج بهتری به نسبت معیار جاکارد ایجاد کرده، هم چنین استفاده از روش عددی بهتر از روش باینری عمل کرده است و دقت و MAP و MRR بیشتری دارد.

بخش سوم

در این بخش بر اساس مدل احتمالاتی BM25 داکيومنت های مشابه هر کویری را حساب می کنیم. این مدل احتمالاتی فرمولاسیون به صورت زیر دارد:

$$\sum_{t \in q} \log \left[\frac{N}{df_t} \right] \cdot \frac{(k_1 + 1)tf_{td}}{k_1((1 - b) + b \times (L_d/L_{ave})) + tf_{td}}$$

این الگوریتم را با سه مقدار b (۰,۵ و ۱ و ۱,۵) و سه مقدار (۰,۲۵ و ۰,۵ و ۰,۷۵) محاسبه می کنیم. حاصل خروجی این بخش برای هر کویری به صورت زیر است:

```
C:\Users\zdehg\AppData\Local\Programs\Python\Python37\python.exe
"D:/univesity/fogelisans/information retrieval/HW1_98131059/TF_IDF.py"
C:\Users\zdehg\AppData\Local\Programs\Python\Python37\lib\site-
packages\sklearn\externals\joblib\externals\cloudpickle\cloudpickle.py:47:
DeprecationWarning: the imp module is deprecated in favour of importlib; see the module's
documentation for alternative uses
import imp
parsing file finished
number of docs: 4600
number of distinct words: 72786
avg length of docs: 355.8517391304348
doc with max length: 1178s1
doc with min length: 3109s1
calculating IDF finished
0
calculating TF-IDF finished
<<<<<<----- part b ----->>>>>>
compute 15 similar with BM25 model
b = 0.5 k = 1.2
6      59395s2
6      51230s8
6      317s1
6      42870s2
6      2546s1
6      16629s1
```

6	5318s1
6	1324s1
6	1100s1
6	433s1
6	2152s1
6	50838s8
6	3313s1
6	441s1
6	2181s1
7	46024s2
7	52667s2
7	51129s4
7	54335s7
7	48207s1
7	45258s6
7	55779s3
7	22286s1
7	1338s1
7	51819s3
7	44728s5
7	2875s1
7	55817s7
7	47225s2
7	45446s8
8	1634s1
8	1961s1
8	51029s5
8	1747s1
8	441s1
8	1918s1
8	524s1
8	50970s5
8	2283s1
8	1215s1
8	865s1
8	59055s7
8	2222s1
8	781s1
8	322s1
9	3105s1
9	1170s1
9	3056s1
9	52667s2
9	322s1
9	54446s2
9	58259s4
9	1650s1
9	2315s1
9	45375s1
9	1297s1
9	2503s1
9	3036s1
9	2544s1
9	368s1
10	3604s1
10	50084s1
10	59051s3
10	47654s10
10	322s1
10	54362s1
10	44244s1
10	2524s1
10	45458s6

```
10 42844s2
10 51034s1
10 53363s1
10 54292s1
10 43389s2
10 41381s2
b = 0.5 k = 1.5
6 59395s2
6 51230s8
6 42870s2
6 1324s1
6 5318s1
6 317s1
6 1100s1
6 2546s1
6 16629s1
6 441s1
6 433s1
6 2181s1
6 3948s1
6 943s1
6 2152s1
7 46024s2
7 52667s2
7 55779s3
7 48207s1
7 54335s7
7 45258s6
7 51129s4
7 22286s1
7 1338s1
7 2875s1
7 44728s5
7 3966s1
7 51819s3
7 44061s2
7 55817s7
8 1634s1
8 1961s1
8 441s1
8 51029s5
8 1747s1
8 1918s1
8 524s1
8 865s1
8 2283s1
8 59055s7
8 1215s1
8 50970s5
8 781s1
8 1178s1
8 322s1
9 3105s1
9 1170s1
9 3056s1
9 52667s2
9 322s1
9 58259s4
9 54446s2
9 1650s1
9 2315s1
9 45375s1
9 2544s1
```

```
9      2503s1
9      1297s1
9      3049s1
9      368s1
10     3604s1
10     50084s1
10     59051s3
10     322s1
10     47654s10
10     44244s1
10     2524s1
10     54362s1
10     51034s1
10     45458s6
10     53363s1
10     42844s2
10     54292s1
10     44902s3
10     43389s2
B = 0.5 k = 2
6      59395s2
6      1324s1
6      42870s2
6      1100s1
6      5318s1
6      51230s8
6      441s1
6      3948s1
6      2546s1
6      317s1
6      943s1
6      433s1
6      16629s1
6      2181s1
6      41820s8
7      55779s3
7      52667s2
7      48207s1
7      46024s2
7      54335s7
7      45258s6
7      2875s1
7      51129s4
7      22286s1
7      1338s1
7      3966s1
7      44728s5
7      2420s1
7      44061s2
7      3551s1
8      1634s1
8      441s1
8      1918s1
8      1961s1
8      51029s5
8      524s1
8      865s1
8      1747s1
8      59055s7
8      2283s1
8      1178s1
8      781s1
8      44195s1
```



```
8      1217s1
8      322s1
9      3105s1
9      3056s1
9      1170s1
9      322s1
9      52667s2
9      58259s4
9      2544s1
9      54446s2
9      3049s1
9      1650s1
9      45375s1
9      2503s1
9      2315s1
9      55779s3
9      3966s1
10     3604s1
10     50084s1
10     322s1
10     59051s3
10     47654s10
10     44244s1
10     51034s1
10     2524s1
10     54362s1
10     53363s1
10     42844s2
10     45458s6
10     54292s1
10     44902s3
10     59682s1
B = 0.75 k = 1.2
6      59395s2
6      51230s8
6      766s1
6      16629s1
6      2546s1
6      42870s2
6      3313s1
6      317s1
6      50838s8
6      2152s1
6      5318s1
6      433s1
6      2074s1
6      2181s1
6      1297s1
7      46024s2
7      51129s4
7      52667s2
7      54335s7
7      45258s6
7      48207s1
7      55779s3
7      52523s3
7      47225s2
7      45446s8
7      51819s3
7      22286s1
7      55817s7
7      50542s5
7      44061s2
```

8 1747s1
8 51029s5
8 1961s1
8 1987s1
8 1634s1
8 441s1
8 50970s5
8 1918s1
8 2222s1
8 1731s1
8 1215s1
8 53s1
8 3533s1
8 48155s1
8 524s1
9 3105s1
9 54446s2
9 1170s1
9 2315s1
9 1650s1
9 3036s1
9 52667s2
9 3056s1
9 1297s1
9 58259s4
9 1389s1
9 322s1
9 47125s5
9 46024s2
9 1142s1
10 59051s3
10 50084s1
10 3604s1
10 53363s1
10 45458s6
10 41381s2
10 42844s2
10 47654s10
10 4017s1
10 43389s2
10 12326s1
10 322s1
10 54362s1
10 2322s1
10 3235s1
B = 0.75 k = 1.5
6 59395s2
6 51230s8
6 16629s1
6 766s1
6 2546s1
6 42870s2
6 3313s1
6 317s1
6 5318s1
6 50838s8
6 2152s1
6 433s1
6 2181s1
6 1297s1
6 2074s1
7 46024s2
7 51129s4

7	52667s2
7	54335s7
7	45258s6
7	55779s3
7	48207s1
7	52523s3
7	47225s2
7	45446s8
7	22286s1
7	51819s3
7	44061s2
7	55817s7
7	50542s5
8	1747s1
8	51029s5
8	1961s1
8	1634s1
8	1987s1
8	441s1
8	1918s1
8	50970s5
8	2222s1
8	1215s1
8	524s1
8	2283s1
8	48155s1
8	59055s7
8	3533s1
9	3105s1
9	1170s1
9	54446s2
9	2315s1
9	1650s1
9	52667s2
9	3036s1
9	3056s1
9	322s1
9	1297s1
9	58259s4
9	1389s1
9	47125s5
9	46024s2
9	1142s1
10	59051s3
10	50084s1
10	3604s1
10	53363s1
10	45458s6
10	42844s2
10	47654s10
10	41381s2
10	43389s2
10	12326s1
10	322s1
10	4017s1
10	54362s1
10	42533s4
10	3235s1
B = 0.75 k = 2	
6	59395s2
6	51230s8
6	16629s1
6	2546s1

6	42870s2
6	5318s1
6	766s1
6	3313s1
6	2181s1
6	433s1
6	317s1
6	50838s8
6	2152s1
6	45702s7
6	1324s1
7	46024s2
7	51129s4
7	52667s2
7	54335s7
7	55779s3
7	45258s6
7	48207s1
7	22286s1
7	52523s3
7	47225s2
7	44061s2
7	45446s8
7	51819s3
7	1338s1
7	55817s7
8	1747s1
8	51029s5
8	1634s1
8	1961s1
8	441s1
8	1918s1
8	50970s5
8	524s1
8	1987s1
8	1215s1
8	2222s1
8	59055s7
8	2283s1
8	865s1
8	3708s1
9	3105s1
9	1170s1
9	54446s2
9	3056s1
9	52667s2
9	2315s1
9	1650s1
9	322s1
9	3036s1
9	58259s4
9	1297s1
9	47125s5
9	46024s2
9	1389s1
9	2865s1
10	59051s3
10	50084s1
10	3604s1
10	53363s1
10	42844s2
10	45458s6
10	47654s10

```
10      322s1
10      41381s2
10      12326s1
10      43389s2
10      54362s1
10      59682s1
10      44902s3
10      44935s1
b = 1 k = 1.2
6       766s1
6       596s1
6       51230s8
6       59395s2
6       16629s1
6       341s1
6       2074s1
6       2977s1
6       2546s1
6       21181s1
6       3313s1
6       50838s8
6       2152s1
6       601s1
6       42870s2
7       46024s2
7       51129s4
7       52523s3
7       50542s5
7       47225s2
7       45446s8
7       45258s6
7       52667s2
7       54335s7
7       9454s1
7       23990s1
7       8762s1
7       48207s1
7       55779s3
7       55817s7
8       1987s1
8       1747s1
8       1731s1
8       51029s5
8       53s1
8       1961s1
8       15232s1
8       2465s1
8       2222s1
8       3533s1
8       50970s5
8       1634s1
8       2538s1
8       415s1
8       48155s1
9       54446s2
9       3036s1
9       3105s1
9       2315s1
9       2413s1
9       407s1
9       1650s1
9       1389s1
9       1170s1
```

```
9      2646s1
9      1921s1
9      2803s1
9      2317s1
9      2310s1
9      47125s5
10     4017s1
10     59051s3
10     1731s1
10     41381s2
10     53363s1
10     50084s1
10     45458s6
10     42844s2
10     4097s1
10     45817s3
10     12326s1
10     46205s2
10     15232s1
10     2322s1
10     43389s2
b = 1 k = 1.5
6      766s1
6      596s1
6      51230s8
6      59395s2
6      16629s1
6      341s1
6      2074s1
6      2977s1
6      2546s1
6      21181s1
6      3313s1
6      50838s8
6      2152s1
6      601s1
6      42870s2
7      46024s2
7      51129s4
7      52523s3
7      50542s5
7      47225s2
7      45446s8
7      45258s6
7      52667s2
7      54335s7
7      9454s1
7      23990s1
7      8762s1
7      48207s1
7      55779s3
7      55817s7
8      1987s1
8      1747s1
8      1731s1
8      51029s5
8      53s1
8      1961s1
8      15232s1
8      2465s1
8      2222s1
8      3533s1
8      50970s5
```

```
8      1634s1
8      2538s1
8      415s1
8      48155s1
9      54446s2
9      3036s1
9      3105s1
9      2315s1
9      2413s1
9      407s1
9      1650s1
9      1389s1
9      1170s1
9      2646s1
9      1921s1
9      2803s1
9      2317s1
9      2310s1
9      47125s5
10     4017s1
10     59051s3
10     1731s1
10     41381s2
10     53363s1
10     50084s1
10     45458s6
10     42844s2
10     4097s1
10     45817s3
10     12326s1
10     46205s2
10     15232s1
10     2322s1
10     43389s2
b = 1 k = 2
6      766s1
6      596s1
6      51230s8
6      59395s2
6      16629s1
6      341s1
6      2074s1
6      2977s1
6      2546s1
6      21181s1
6      3313s1
6      50838s8
6      2152s1
6      601s1
6      42870s2
7      46024s2
7      51129s4
7      52523s3
7      50542s5
7      47225s2
7      45446s8
7      45258s6
7      52667s2
7      54335s7
7      9454s1
7      23990s1
7      8762s1
7      48207s1
```

```
7 55779s3
7 55817s7
8 1987s1
8 1747s1
8 1731s1
8 51029s5
8 53s1
8 1961s1
8 15232s1
8 2465s1
8 2222s1
8 3533s1
8 50970s5
8 1634s1
8 2538s1
8 415s1
8 48155s1
9 54446s2
9 3036s1
9 3105s1
9 2315s1
9 2413s1
9 407s1
9 1650s1
9 1389s1
9 1170s1
9 2646s1
9 1921s1
9 2803s1
9 2317s1
9 2310s1
9 47125s5
10 4017s1
10 59051s3
10 1731s1
10 41381s2
10 53363s1
10 50084s1
10 45458s6
10 42844s2
10 4097s1
10 45817s3
10 12326s1
10 46205s2
10 15232s1
10 2322s1
10 43389s2
```

Process finished with exit code 0

با توجه به نتایج بدست آمده در بخش چهارم، بهترین نتیجه با استفاده از مقدار $k = 0.75$ $b = 0.2189$ با مقدار MAP، 0.2189 است. در این بخش با افزایش مقدار k به طور کلی روند خاص پر رنگی وجود ندارد اما میزان معیار MAP و MRR تا حد کمی افزایش می یابد. بیشترین مقدار MAP و MRR نیز در b برابر 0.75 بدست آمده و مقادیر دیگر معیار های تقریبا برابر و کمتری دارند.

بخش چهارم)

در این بخش نتایج بخش های قبل را با داده judgment با هر ۴ معیار خواسته شده، مقایسه می کنیم:

```
C:\Users\zdehg\AppData\Local\Programs\Python\Python37\python.exe
"D:/univesity/foqelisans/information retrieval/HW1_98131059/TF_IDF.py"
C:\Users\zdehg\AppData\Local\Programs\Python\Python37\lib\site-
packages\sklearn\externals\joblib\externals\cloudpickle\cloudpickle.py:47:
DeprecationWarning: the imp module is deprecated in favour of importlib; see
the module's documentation for alternative uses
    import imp
<<<<<<----- part c ----->>>>>>
evaluation metrics : p@5 p@10 MRR MAP
compute metrics for result of part a :
analyse result of part A :
<<----- p@5 ----->>
numeric cosine : [0.3333, 0.4667, 0.9333, 0.3333, 0.9333]
numeric jaccard : [0.0, 0.0, 0.0, 0.0, 0.0]
binary cosine : [0.2, 0.4, 0.0, 0.0667, 0.7333]
binary jaccard : [0.0667, 0.2, 0.0, 0.0, 0.6667]
<<----- p@10 ----->>
numeric cosine : [0.4, 0.4, 1.0, 0.3, 1.0]
numeric jaccard : [0.0, 0.0, 0.0, 0.0, 0.0]
binary cosine : [0.2, 0.5, 0.0, 0.1, 0.6]
binary jaccard : [0.1, 0.2, 0.0, 0.0, 0.6]
<<----- MAP ----->>
numeric cosine : 0.5865
numeric jaccard : 0.0
binary cosine : 0.2766
binary jaccard : 0.2478
<<----- MRR ----->>
numeric cosine : 0.6167
numeric jaccard : 0.0
binary cosine : 0.465
binary jaccard : 0.4222

analyse result of part B :
<<----- p@5 ----->>
b = 0.50 k = 1.2 : [0.0667, 0.6667, 0.0667, 0.0, 0.2667]
b = 0.50 k = 1.5 : [0.0667, 0.6, 0.0667, 0.0, 0.2667]
b = 0.50 k = 2.0 : [0.0667, 0.4667, 0.0667, 0.0, 0.2]
b = 0.75 k = 1.2 : [0.0667, 0.6667, 0.0667, 0.0, 0.3333]
b = 0.75 k = 1.5 : [0.0667, 0.6667, 0.0667, 0.0, 0.3333]
b = 0.75 k = 2.0 : [0.0667, 0.6667, 0.0667, 0.0, 0.4]
b = 1.00 k = 1.2 : [0.0667, 0.6, 0.0667, 0.0, 0.4667]
b = 1.00 k = 1.5 : [0.0667, 0.6, 0.0667, 0.0, 0.4667]
b = 1.00 k = 2.0 : [0.0667, 0.6, 0.0667, 0.0, 0.4667]
<<----- p@10 ----->>
b = 0.50 k = 1.2 : [0.1, 0.6, 0.1, 0.0, 0.2]
b = 0.50 k = 1.5 : [0.1, 0.5, 0.0, 0.0, 0.1]
```

```
b = 0.50 k = 2.0 : [0.1, 0.5, 0.0, 0.0, 0.1]
b = 0.75 k = 1.2 : [0.0, 0.7, 0.1, 0.0, 0.4]
b = 0.75 k = 1.5 : [0.1, 0.7, 0.1, 0.0, 0.5]
b = 0.75 k = 2.0 : [0.1, 0.6, 0.1, 0.0, 0.4]
b = 1.00 k = 1.2 : [0.1, 0.7, 0.0, 0.0, 0.4]
b = 1.00 k = 1.5 : [0.1, 0.7, 0.0, 0.0, 0.4]
b = 1.00 k = 2.0 : [0.1, 0.7, 0.0, 0.0, 0.4]
```

```
<<----- MAP ----->>
```

```
b = 0.50 k = 1.2 : 0.1827
b = 0.50 k = 1.5 : 0.165
b = 0.50 k = 2.0 : 0.1557
b = 0.75 k = 1.2 : 0.2151
b = 0.75 k = 1.5 : 0.2184
b = 0.75 k = 2.0 : 0.2189
b = 1.00 k = 1.2 : 0.1897
b = 1.00 k = 1.5 : 0.1897
b = 1.00 k = 2.0 : 0.1897
```

```
<<----- MRR ----->>
```

```
b = 0.50 k = 1.2 : 0.1758
b = 0.50 k = 1.5 : 0.1767
b = 0.50 k = 2.0 : 0.1754
b = 0.75 k = 1.2 : 0.1968
b = 0.75 k = 1.5 : 0.1972
b = 0.75 k = 2.0 : 0.2119
b = 1.00 k = 1.2 : 0.1782
b = 1.00 k = 1.5 : 0.1782
b = 1.00 k = 2.0 : 0.1782
```

```
Process finished with exit code 0
```