Kendra Maggiore

Homework 4

Spring 2020

1.

1.1:

|  |
| --- |
| Forecast -> 1 |
| Home -> 1,2,3,4 |
| In -> 2,3 |
| Increase -> 3 |
| July -> 2,3,4 |
| New -> 1,4 |
| Raise -> 2,4 |
| Sale -> 1,2,3,4 |
| Top -> 1 |

1.2:

It would be documents 1 and 3 because Gates appears within 2 words of Microsoft

2.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | query | | | | document | | | |
| terms | df | idf | tf | log tf | q | q’ | tf | log tf | d | d’ |
| bad | 1000 | 4 | 0 | 0 | 0 | 0 | 2 | 1.3 | 1.3 | 0.56 |
| good | 10000 | 3 | 1 | 1 | 3 | 0.79 | 2 | 1.3 | 1.3 | 0.56 |
| instructor | 10 | 6 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0.43 |
| student | 50000 | 2.3 | 1 | 1 | 2.3 | 0.61 | 1 | 1 | 1 | 0.43 |

The cosine similarity is 0.70

3.

3.1:

P =8/8+10 = 0.44

R = 8/8+12 = 0.40

F balanced = 1 / ½(1/0.44)+(1/0.4) = 0.42

3.2:

A. System 1 MAP = (1.0+0.66+0.33+0.4)/4 = 0.6

System 2 MAP = (0.5+0.4+0.5+0.57)/4 = 0.49

System 1 has a higher MAP

B.

System 1 R-Precision = 2/4 = ½

System 2 R-Precision = ¼

Yes, it does rank the systems the same as MAP

3.3:

A. P = 6/20 = 0.3

B. F balanced = 1/½ (1/0.3 + 1/0.75) = 0.43

C. 8\*0.25= 2 Then use it for the first 8 points:

2/2 ,2/3,2/4,2/5.2/6,2/6,2/7.2/8

D. Since there are 11 points it would be 4/11=0.36

E. MAP = (1+1+0.3+0.36+0.3+0.3)/6 = 0.54

F. The largest possible MAP would be if all 8 relevant documents where as close as they could be. So, 1/8(1+1+0.3+0.36+0.3+0.3+0.3+0.36) = 0.50

G. The smallest possible MAP would be if the last 2 relevant documents where far away from the first six. So, 1/8(1+1+0.3+0.36+0.3+0.3+0.0007+0.0008) = 0.41

4.1:

A. 0 1 0

0.5 0 0.5

0 1 0

B. N = 3

1/6 2/3 1/6

5/12 1/6 5/12

1/6 2/3 1/6

C. N= 3

1/3 1/3 1/3

1/3 1/3 1/3

1/3 1/3 1/3

4.2:

Page Rank:

0.9(0 0.5 0.5) + 0.1(1/3 1/3 1/3)

Q1 = 1/30

Q2 = 29/60

Q3 = 29/60

Authority (in to Node):

Q1 = 0

Q2=1

Q3 =1

Hub (out of node) Scores after normalized:

Q1 = 1

Q2= 0.5

Q3=0.5