

יש להחזיר את השאלון – יש לענות על שאלה 5 על גבי הגיליון.

- מסמכים רלוונטיים ולשאלת q_2 8 מסמכים רלוונטיים.

$$\rightarrow \text{recall} = \frac{8}{8} = 1$$

- הנח שלצורך חישוב הווקטור של השאילתא והמסמכים משתמשים רק ב tf מנורמל לאורך המסמך – ללא idf. מה תהיה השאילתא המעודכנת לאחר הפעלת אלגוריתם rocchio אם משקל של השאילתא הנוכחית הוא 1, משקל של מסמך רלוונטי 0.75, משקל מסמך לא רלוונטי 0.25.

5. 14%

א. (6%) נתונה שאילתא q1 : A and B and C
נתון מסמך d1 המכיל את ה - A Terms ו-B
ומסמך d2 שלא מכיל אף אחד מה Terms הנ"ל.

סמן נכון/לא נכון ליד כל משפט

1. 2% מודל בוליאני טהור לא יחזיר אף אחד משני המסמכים, d1, d2 לשאילתא q1 נכון/לא

נכון

2. 2% כל מודל בוליאני מורחב יחזיר רק את d1. נכון/לא נכון

3. 2% המודל הווקטורי לא תומך בשאילתות כמו q1. נכון/לא נכון

ב. 8% נתונה שאילתא q1 : C,B,A

נתון המאגר הבא: המספרים מסמלים תדירות של ה Terms במסמך.

מסמך d1 הכולל את ה terms A (3), B (5), C (10)

מסמך d2 הכולל את, B (1), C(5)

מסמך d3 הכולל את A(30)

מסמך d4 הכולל את C(5), B(2)

מסמך d5 הכולל את C(5), B(2)

מסמך d6 שלא כולל אף אחד מה terms הנ"ל

סמן נכון/לא נכון ליד כל משפט, בהנחה שמשתמשים בשיטת שקלול של tf*idf

1. 2% ייתכן שמסמך d6 יוחזר לשאילתא q1 אם יופעל אלגוריתם lsi והמסמך קשור לנושא

השאילתא. נכון/לא נכון

2. 2%, המודל הווקטורי משתמש ב inner-product ידרג את d1 במקום הראשון.

נכון/לא נכון

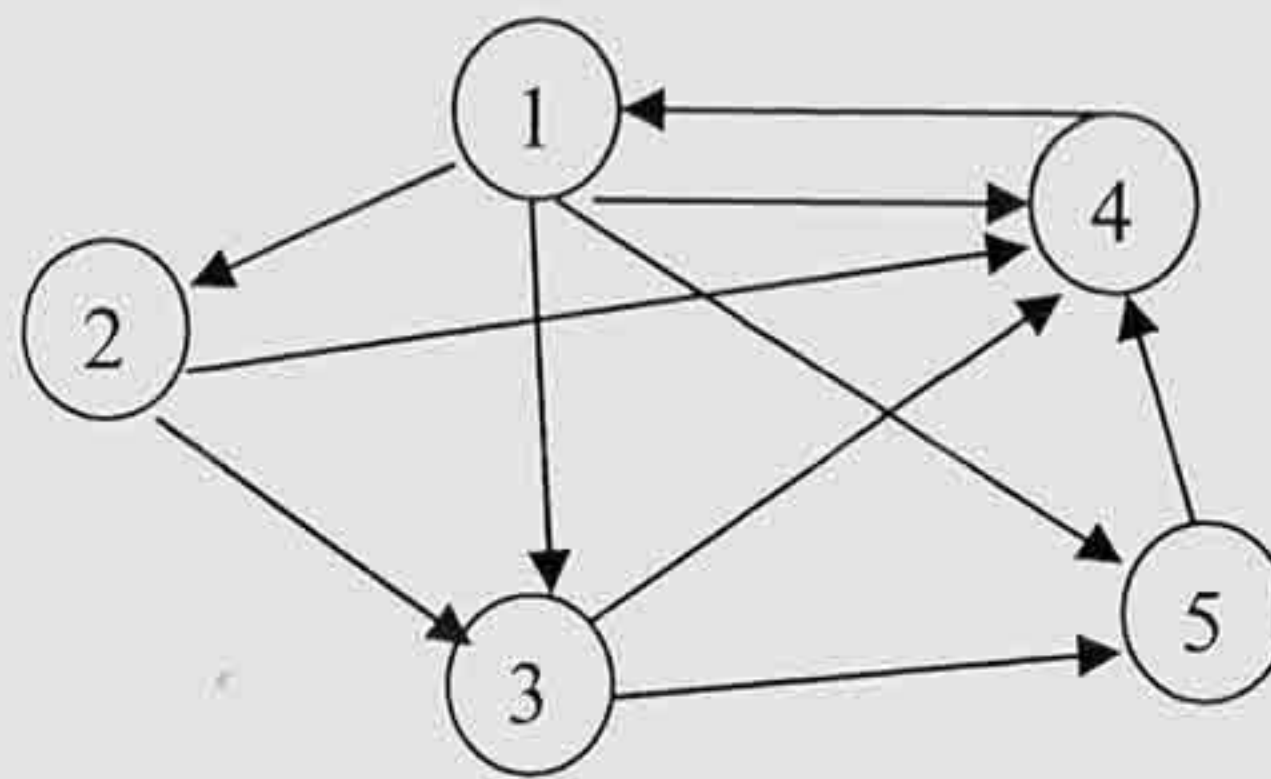
3. 2% הוספה של ה A Term למסמך d2 תשפיע רק על רמת הדמיון בין d2 לשאילתא.

נכון/לא נכון

4. 2% הוספה של ה A term למסמך d4 תשפיע בין היתר על רמת הדמיון בין מסמך d1

לשאילתא. נכון/לא נכון

6. 26% נתונה הרשת הבאה:



- א. 10% מיהו הצומת/צמתים בעל ה pagerank הגבוה ביותר, הצומת בעל ה pagerank הנמוך ביותר
- ב. 8% מיהו הצומת בעל ה hub הגבוה ביותר וה authority הנמוך ביותר.
- ג. 8% מה יקרה ל pagerank של צומת 1 אם נבטל את הקישור בין צומת 5 לצומת 4.
7. 10% כדי לעשות pagerank מותאם למשתמש אפשר להתאים את הפרמטר d בנוסחא בהתאם לתחומי העניין של המשתמש, כלומר לאתרים שהם בתחום העניין של המשתמש אפשר להגדיל את ההסתברות של המשתמש להגיע אליהם,
- א. 4% מה הבעיה היישומית של pagerank מותאם אישית?
- ב. 6% הצע פתרון יישומי (באופן כללי, אין צורך לכתוב אלגוריתם או נוסחאות, רק הסבר).

בהצלחה
ברכה ואיליה

$$\text{idf} = \log_2 \frac{N}{df_i}$$

(1)

מכיון שה-term נחשב בכל המסמכים במאגר
 נקרא $N = df_i$ וכן $\text{idf} = 0$

עקרון של נדירות קו חישוב term-ה- stop words
 חישוב ההצמדה של המסמך לשאלות.
 מכיון idf של היה אם לא ההכלה
 כי לא ישב מחזור ההצמדה לשאלות.

יחידות ההכלה: stop words
 הקלטה האנציקס וקובץ posting, כי לא
 נכללים אוגו עץ.

מחישוב החישוב תשפיר, מכיון שזרז פחות
 חישובים.

מחישוב החישוב גשפיר, מכיון שזרז לחסם פחות
 terms האנציקס ולפי האנציקס עצמו היה קלן ואור.



2

אל מוצא שר הסכום אל
מסר אל דוק התוצאות.

מכיוון שמעמיק האגן שילוב בירור
המסמכים ואל התנאים להדבס
מסמכים להיק.

תקרה מסוים שבו כן יכול
להיות סיפור זה אל קרפי

ה posting מהמלך השונה
יבוא אל שרת שונים ואל השרת
של הקבצים האית"ם יפול
והשרת של הקבצים הפחית אוביק
יאלר לפסול אל נקבה תוצאות
כפשוט ואל אל אל הקבצים היו

יושבים אל שרת אחר ואל היתה תוקה ערמול והשרת
היה נוסף אל אל היינו מקבלים תוצאות כולל.
אלא מעניין יקם אל המעלה מתאזי אל מסמכים לפעולתם
יש סיבוי אל טיפא ואל ילתי מסמכים כמותיים קצתים.

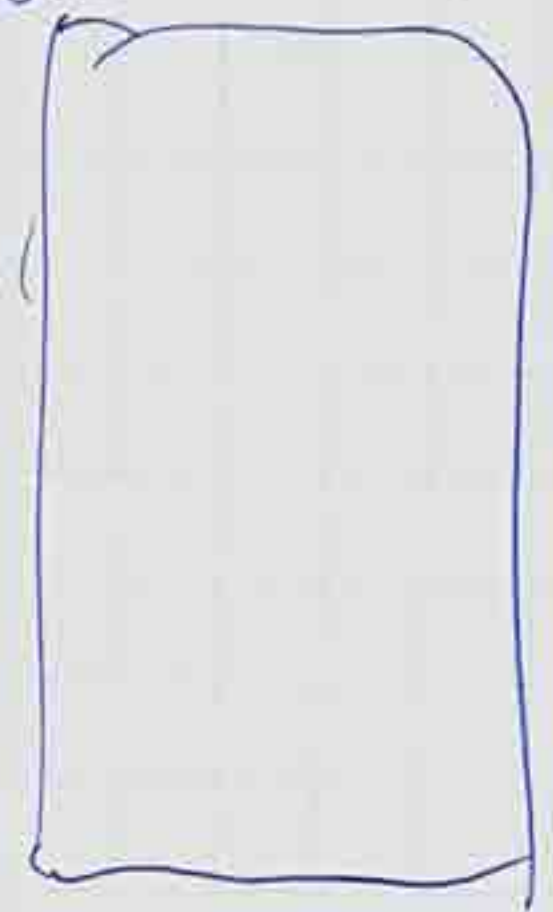
ייתכן גם מכיוון שקודם כל מחשבים ברמה
ההולנדית ואל יש גם מסמכים תוצאות אל מחלוקת
אגן לכן החיפוש קאן מכיוון שזכר לעצב
פחית מצב.

ייתכן גם 2 ש למשל עבור שאלתא ש AND בין
ה terms אל עבור אחר ה terms אל נמצא
מסמכים בסכך ההולנדית אל אשר עובד אל
החיפוש קבלה ואל בורק עבור אל
ה terms ליה אל נכחן! אולי יש בעיה השמלה
מסמכים מתחלפות

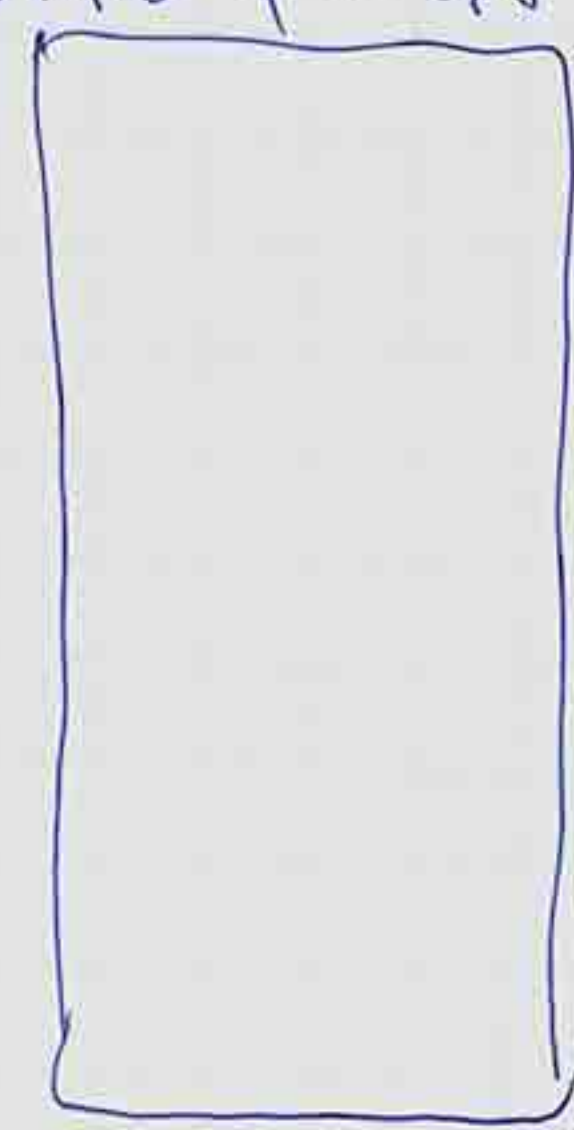
3

אגן מרכזי

מסמכים
אית"ם



מסמכים פחית אית"ם



3

recall_{q₁} = $\frac{6}{12} = 0.5$
recall_{q₂} = $\frac{8}{8} = 1$

מסמך 10000
נדרש 128
נדרש 128

Precision_{q₁} = $\frac{6}{20} = \frac{3}{10} = 0.3$ ✓

F-measure_{q₁} = $\frac{2PR}{P+R} = \frac{2 \cdot 0.3 \cdot 0.5}{0.3+0.5} = 0.375$ ✓

25% recall = $0.25 = \frac{3}{12}$ (2)

~~P = 0.25~~
~~R = 0.25~~
~~F = 0.25~~
 $P_{0.25} = \frac{3}{9} = \frac{1}{3}$ ✓

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

	Q ₁	recall	pre	Q ₂	rec	pre
1	✓	1/12 = 0.08	1	X		
2	✓	2/12 = 0.16	1	X		
3	X			✓	1/8 (0.125)	1/3
4	X			✓	2/8 0.25	2/4 = 1/2
5	X			X		
6	X			✓	3/8 0.375	3/6 = 1/2
7	X			✓	4/8 0.5	4/7
8	X			X		
9	✓	3/12 = 0.25	3/9 = 1/3	✓	5/8 0.625	5/9
10	X			X		
11	✓	4/12 = 0.33	4/11	X		
12	X			X		
13	X			X		
14	X			✓	6/8 0.75	6/14 = 3/7
15	✓	5/12 = 0.41	5/15 = 1/3	✓	7/8 0.875	7/15
16	X			✓	8/8 1	8/16 = 1/2
17	X			X		
18	X			X		
19	X			X		
20	✓	6/12 = 0.5	6/20 = 3/10	X		

או האק לה סה"כ הוליוואיג שאלו.
 עפ"י Q_2 זה א משנה (8 דף מקרה)
 אדם עפ"י Q_1 בחינת עקדת אל ס (סה"כ הוליוואיג)
 שאלו ולא נ"ל. דף אופן יס"ל עפ"י שהיה צריך
 עמלק ד נ"ל, אלא החילוק בולט.

4

$$\alpha = 1$$

$$\beta = 0.75$$

$$\gamma = 0.25$$

<u>like</u>	cheap	CD's	DVD's	extremely
$\frac{1}{7}$	$\frac{3}{7}$	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{1}{7}$

$$Q_0 = \left(\frac{3}{7}, \frac{2}{7}, \frac{1}{7}, \frac{1}{7} \right)$$

$$d_1 = \left(\frac{2}{5}, \frac{2}{5}, 0, 0 \right)$$

$$d_2 = \left(\frac{1}{3}, 0, \frac{1}{3}, 0 \right)$$

האלימות
האלימות
האלימות
האלימות

$$\begin{aligned} \vec{q}_m &= \left(\frac{3}{7}, \frac{2}{7}, \frac{1}{7}, \frac{1}{7} \right) + 0.75 \cdot \left(\frac{2}{5}, \frac{2}{5}, 0, 0 \right) - 0.25 \left(\frac{1}{3}, 0, \frac{1}{3}, 0 \right) \\ &= (0.65, 0.59, 0.06, 0.14) \end{aligned}$$

האלימות
האלימות
האלימות
האלימות



software, thrills

5) (c) $q_{r,p}$ A and B and C $(1,1,1)$

~~Handwritten scribble~~

1) $|1\rangle$

2) $|1\rangle$

3) $|1\rangle$ $|1\rangle$

2)

$q_{r,p}$	A	B	C
$id \neq$	1	1	1
d_1	-	-	-
d_2	-	-	-
d_3	-	-	-
d_4	-	-	-
d_5	-	-	-
d_6	X	0	0

6

ניבן ליכאן עצומה 4 חקנה אל כס
ה rank 5, חצי מה rank
3, חצי מה rank 2
1 חצי מה rank 1

פסן היא הצומה שהיה לה הציון

הציון ביותר. (-1) (-2)

✓ הצומה בעל הציון הנמוך ביותר
היא 2, מכיון שהיא מקבלת רק $\frac{1}{4}$ מהציון
של 1.

עצומה צומה 3 מקבל בנוסף לקצק
ה rank 2.
צומה 5 מקבל בנוסף ל- $\frac{1}{4}$ של 1
חצי חצי מה rank 3.

✓

הצומה קבלה הטבה ביותר הוא (1), מכיון
שזוהי מפתח 4 צומות אחרים ואחד מהם
הוא (4) שיש לו Authority גבוהה.

הצומה 8 Authority הנמוך ביותר היא
(2) מכיון שהיא מקבלת רק $\frac{1}{4}$ מההטבה של צומה (1).
ל-1 (-4)

ה page rank של צומה 1 יתאם מכיון
ש 5 יהיו Dead End. rank 1
~~הצומה 1 מקבלת 0~~
היא 0.

✓

(7)

הקציה הישומית היא ש(צ)ר לשמור
rank שונה עם הצבים ברשת עם משתמש
מכיוון rank אצד אתר מסוים משפיע על אתרים
אתרים השפיע זה (צ)ר עוקים מאלה עצום
ש מ'צ'צ. עם קצמן חלקי למיצר המשתמש ✓
(צ)ר לשמור את א בהתאם לתחומי ענין חצבים
ואז בקצם (צ)ר לחשב מחצב את rank - ?
כל העסק באינדקס של הישגתם הלה מחצב.

(7) אפשר להצדיר כמה גממי ענין כללים
נשמור מאלה (אינדקס) משותף עלכולם.
בנוסף אלו נוכ לשמור עבור כל משתמש
עם אינדקס משלו לפי תחומי הענין שלו.
אם וקם להיות משתמש לא מוכר במערכת אל ✓
נחזיר לו תוצאות מהאינדקס הילי בלבד.
אך כמו משתמש מוכר במערכת והוא מקצד שאלת
קשורה לאחד מתחומי הענין שלו, אז נחפש באינדקס
תחומי הענין עסי התחומים שיוצרים ענין שיהיונליים
למשתמש.

אנליזה מסתמית - דגם רגרסיה

אנחנו מלחמים - נגד המלחמה

