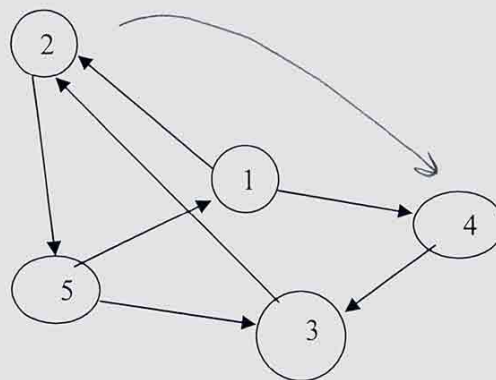


111 70 36 1 8/8

בס"ד

איחזור מידע תש"ע – 372.1.4406 ✓  
 סמסטר חורף מועד א' 31.01.10 ✓  
 ד"ר ברכה שפירא, איליה פרידמן ✓

משך המבחן : שתיים וחצי ✓  
 חומר עזר: מותר (לא מחשב נייד) ✓  
 יש להחזיר את השאלון – נא לענות על שאלות 5-8 על השאלון ✓



1. 25% ✓

1 < 3 < 2  
 5 < 2      1 < 3

- א. (5%) אם נוסף קישור מצומת 2 לצומת 4, האם וכיצד יושפעו ערכי HUB ו - AUTHORITY (על פי אלגוריתם HITS) של צומת 5 (כלומר, האם הערכים יהיו גבוהים, או נמוכים יותר, או ללא שינוי ומדוע). ✓
- ב. (5%) הראה את מטריצת המעברים (לצורך חישוב pagerank) של הרשת מסעיף א (לאחר ההוספה של הקישור מסעיף א). ✓
- ג. (5%) חשב שתי איטרציות של pagerank של הגרף שבציור לאחר ההוספה מסעיף א  $d=0.8$ . (ללא נרמול) (לא בשיטת המטריצות). אין צורך לחלק את  $(1-d)$  במספר הצמתים ברשת. את האיטרציה השנייה אין צורך לחשב מספרית, אלא רק להציב ערכים מתאימים בנוסחה. ✓
- ד. (5%) האם אפשר להעריך בגף הנתון בשאלה (לאחר ההוספה ב - א') איזה צומת תהיה בעלת ערך pagerank גבוה ביותר לאחר ההתכנסות. אם כן הערך והסבר את הערכתך, או הסבר מדוע אי אפשר להעריך. ✓
- ה. (5%) האם אלגוריתמים של link-analysis כדוגמת pagerank מודדים גם את פופולאריות הדף אצל הגולשים. אם כן, כיצד? ואם לא איך אפשר למדוד פופולאריות של דפים אצל גולשים? ✓

2. 8% - רשומה בקובץ posting הופכי כוללת נתונים על פי terms. על כל term הקובץ כולל את רשימת כל המסמכים שבהם הוא מופיע, ואינפורמציה נוספת (למשל, מיקום המופעים של ה term במסמכים). רשימת המסמכים לכל term בדרך כלל ממוינת באחת משתי הדרכים הבאות: ✓
- א. מיון על פי זיהוי המסמך. ✓
- ב. מיון על פי שכיחות (חשיבות, משקל), המילים במסמכים, כלומר מסמכים שבהם ה-term יותר משמעותי יופיעו ברשימה לפני מסמכים שבהם ה term פחות משמעותי. ✓
- ציין יתרון אחד לכל אחת מהשיטות.

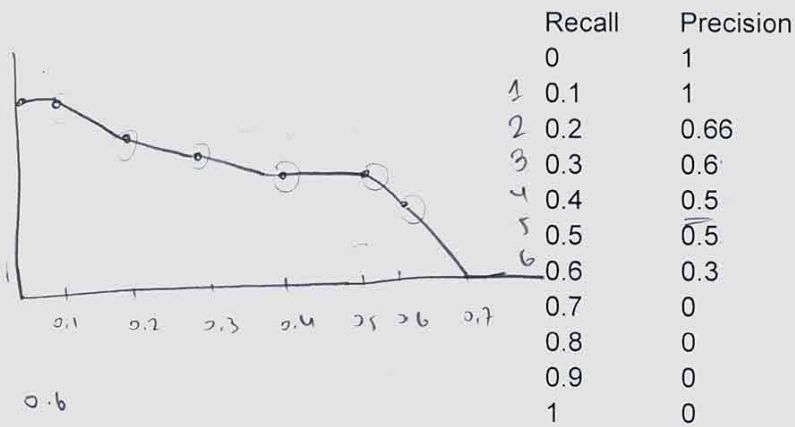
1 0.312  
 2 0.664  
 3 0.6  
 4 0.474  
 5 0.216



מ 70 36 1 8/8

$$\frac{2}{x} = 0.6$$

7. 14% לשאילתא מסוימת התקבלו הערכים בגרף P-R (עם אינטרפולציה) כאשר המנוע החזיר 20 מסמכים וישנם סך הכול במאגר 10 מסמכים רלוונטיים לשאילתא. להלן הערכים בגרף.



מכדול ניון  
לפיכך שפירו  
סך 6 מסמכים  
רלוונטיים

$$5 \rightarrow 0.6$$

$$8 \rightarrow 0.5$$

$$70 \rightarrow \frac{6}{20} = 0.3$$

9% מה היה ה precision לאחר שהמערכת החזירה 5 מסמכים, 8 מסמכים ו-20 מסמכים,  $P=0.3$   
5% מה היה ה precision לאחר שהמערכת החזירה 5 מסמכים רלוונטיים.  $P=0.5$

8. 10% בטבלה הבאה נתונים דירוגים של משתמשים על מוצרים שרכשו (1-5). לצורך ניבוי מוצר למשתמש בשיטה השיתופית לוקחים בחשבון רק שכנים שהדמיון שלהם למשתמש הוא גדול או שווה ל-0.5. לצורך ניבוי ההתאמה של מוצר 5 למשתמש 3, השכנים שיילקחו בחשבון הם (רק תשובה אחת נכונה):

	משתמש 1	משתמש 2	משתמש 3	משתמש 4	משתמש 5
מוצר 1	5	2	5		4
מוצר 2				5	
מוצר 3				5	
מוצר 4	3	5	2		3
מוצר 5		5		5	5

- א. אין שכנים מתאימים  
ב. משתמשים 2 ו-5  
ג. משתמשים 1 ו-5  
ד. משתמש 5  
ה. אף תשובה לא נכונה

משתמש 2 יור  
כחוק ממנו  
באק 0.8

הכי קרוב יור  
(2-3)  
(5-4)

בהצלחה

ברכה ואיליה



היחס הקשר ה- hub של 2 הצדדים שצויי שטחי יחידים וכן

$h(2) \uparrow \rightarrow a(1) \uparrow$  2 le hub-f meo puz af2y, A(1), 5 le kashuta ✓

התוצאה של תהליך כיוון  
שמישהו נ- A של 3 וזהו כזה שיהיה  
מכונה א-  
1 2 3 4 5  
1 2 3 4 5

$$M = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

מטבח	מלכים
מלכות	מלכות
מלכות	מלכות
מלכות	מלכות

6. חתונה      דמי חתונה      1

7. הולדתו

$$d = 0.8$$
$$1 - d = 0.2$$

① נותן ב 3 ציורים -  
 שטח קטן (הקטנה) -  
 (אם נתון זרוע במחצה)  
 שטח קטנה  
 $0.2 = \frac{1}{15}$   
 זרוע קטנה - יחסית  
 (ע)  $u-d$

$$PR^*(p) = d \cdot \sum_{q \in p} \frac{PR(q)}{n_q} + (1-d)$$

$$PR(1) = 0.8 \cdot \frac{0.2}{2} + 0.2 = 0.28$$

$$PR(2) = 0.8 \cdot \left( \overset{1}{\frac{0.2}{2}} + \overset{3}{0.2} \right) + 0.2 = 0.44$$

$$PR(3) = 0.8 \left( 0.2 + \frac{0.2}{2} \right) + 0.2 = 0.44$$

$$PR(4) = 0.8 \left( \frac{0.2}{2} + \frac{0.2}{2} \right) + 0.2 = 0.36$$

$$PR(5) = 0.8 \left( \frac{0.2}{2} \right) + 0.2 = 0.28$$

$$PR(1) = 0.8 \left( \frac{0.28}{2} \right) + 0.2 =$$

$$PR(2) = 0.8 \left( \frac{0.28}{2} + 0.44 \right) + 0.2$$

$$P(13) = 0.8 \left( 0.36 + \frac{0.21}{2} \right) + 0.2$$

$$P_R(4) = 0.8 \left( \frac{0.40}{2} + \frac{0.20}{2} \right) + 0.2$$

$$PR(5) = 0.2 \left( \frac{0.44}{2} \right) = 0.2$$

17350.1c

203010

④ 110 300/1000

3. דרגת כוח של דף PageRank דורש 2.

השינוי 2, 1, 3, 2 מקבל את 3 ודג' חלק (דג' 4) דף 5) חלק 1) את 6

השינוי של דרגת 3, 1, 4 < 3 מקבל את 6 ודג' חלק 4

4 < 3 < 2

5 < 2

1 < 5 < 2

השינוי 2, 1, 5 מקבל את 3 ודג' חלק 2) חלק 1) את 5

השינוי 3, 1, 4 < 3 מקבל את 6 ודג' חלק 4

ה. אדורתיס Link Analysis לא מודד כמות ישיר את הפוטנציאל של דף ויצא החישוב

אלו את החשיבה שלו בין דברים קונקרטיים, קישורים, דג' 5, PR, Authority, PageRank

כדי למדוד באופן ישיר פוטנציאל של דפים יש להשתמש במספר clicks -

מספר כניסות אל דף, פירוק אלפים אדורתיס, המתייחסים דיוקן ישיר וסמנטי

הדג' או אדורתיס, המתייחסים, פוטנציאל דף דג' 5, PR, Authority, PageRank

הנטישה דג' 5, PR, Authority, PageRank, המתייחסים, פוטנציאל דף דג' 5, PR, Authority, PageRank

אחרים דאמן דג' 5, PR, Authority, PageRank

חלק המספר 2 של דף PR דג' 5, PR, Authority, PageRank

דג' 5, PR, Authority, PageRank

שטח כו חמש חטים מחר יאר בוינדקס וואס ביצאם טאגס יאר

הדיונים. טענתו של כץ. מילן. השתתפה, נולד אבד חברים של קצבים

[illegible]

בדקה 20 דמיך קליטת (הנחה) AND דמיך זמן זמן זמן

And while it's true that

מחבר: שמואל א. טל  
מחבר: משה

התורה שכל - term יורה נסבה מן המצויים - המצויים חיוניים.

הכלל האחרון - term שנקרא קטור סדרה והתמנהגים האחרים

שם יחידיו זכר (2) - דאזיגה ירוי דיי חרטים

[illegible]

התנאי הזה לא יחול על מי שיש לו

שם המדינה: ישראל שם העיר: ירושלים שם הרחוב: רחוב המלך דוד

(י. שרג) זקוק , ס. 307 ככה על posthy יפה (אראל) preusm כלומר

המנכ"ל      הכי      האונט"ם      יומל"ו      רג"ח      זמ"ה      כיון      שח"ח      ד"ר

307 22

[illegible][illegible]

✓



מיון נכבד אנו - הנקודה שתואר את שורה הדורשים וזמן תחילת נכבד אנו

השקל של כל דברים כפי אנו אנו יושב H הנוסחה וזמן אנו

דניה נכבד רביון והחוסר אנו אנו רביון נכבד אנו אנו

כפי אנו אנו דברים נכבד אנו אנו דברים נכבד (כיון שזמן דברים אנו אנו)  
אנו אנו אנו אנו אנו אנו

הניסוי שנקבד - נח אנו אנו אנו אנו אנו אנו

אנו 2 אנו אנו אנו אנו אנו אנו

אנו אנו אנו אנו אנו אנו

אנו אנו אנו אנו אנו אנו

אנו אנו אנו אנו אנו אנו

אנו אנו אנו אנו אנו אנו

אנו אנו אנו אנו אנו אנו

אנו אנו אנו אנו אנו אנו

אנו אנו אנו אנו אנו אנו

אנו אנו אנו אנו אנו אנו

$$W_{term} = 2 \cdot \max \left( \begin{array}{l} \text{השקל הדברים ביותר אנו} \\ \text{הניסוי שנקבד רביון} \end{array} \right)$$

הזיון - אנו אנו אנו אנו אנו אנו

$$W_{term} = 2$$

אנו אנו אנו אנו אנו אנו

$$W_{term} = \lambda \cdot \frac{f_{ij}}{\max(f_j)} + \beta \cdot \frac{\text{מספר דברים} \cdot \text{השקל הדברים} \cdot \text{הניסוי שנקבד}}{\text{מספר דברים} \cdot \text{השקל הדברים} \cdot \text{הניסוי שנקבד}}$$

אנו אנו אנו אנו אנו אנו