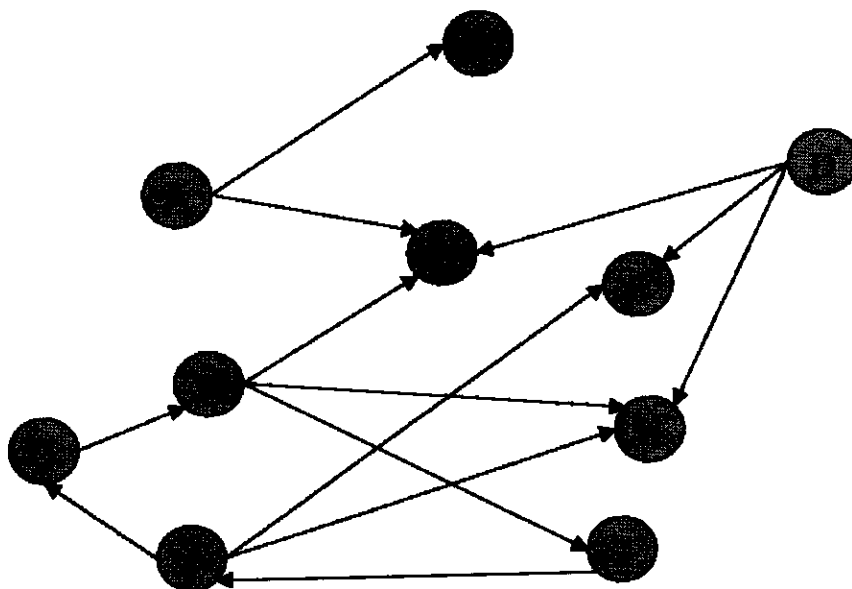


אוניברסיטת בן-גוריון - המחלקה להנדסת מערכות מידע
קורס איחזור מידע וספריות דיגיטליות
סמסטר אביב תשס"ו - 28.07.06 – מועד ב

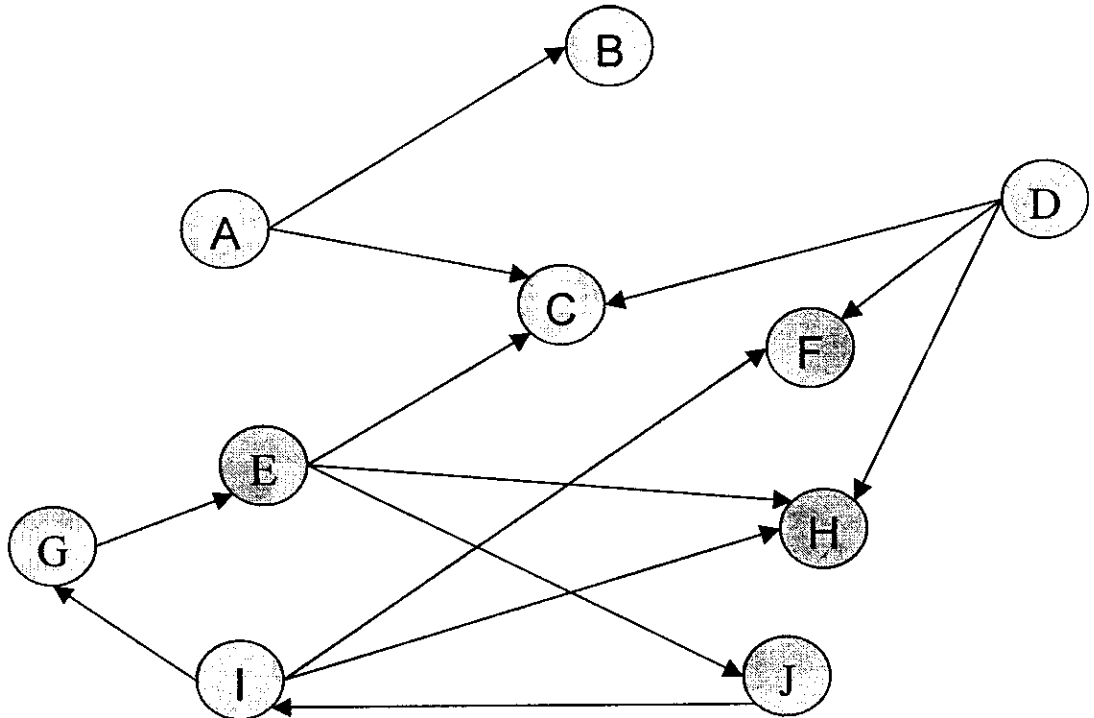
מרצה: ד"ר ברכה שפירא, אסיסטנט : ארז שלום
 משך המבחן: 3 שעות, חומר עזר מודפס או כתוב - מותר, מחשבון- מותר
 יש לענות על כל 5 השאלות. יש להחזיר את השאלונים.

1. (26) שני מנועי חיפוש החזירו תוצאות לאותה שאילתא. כל מנוע החזיר 20 מסמכים. להלן רשימת המסמכים הרלוונטים (מתוך אלו שהוחזרו) לכל מנוע. המספרים מסמלים את סדר ההחזרה, למשל 1 מסמן מסמך רלוונטי שהוחזר במקום הראשון:
 מנוע א': 1, 2, 4, 7, 10
 מנוע ב': 7, 8, 9, 12, 14, 16, 19, 20
 ידוע שהמאגר כולל 8 מסמכים רלוונטים לשאילתא.
 א. (5) חשב לכל מנוע:
 precision-recall כלליים, precision, r-precision בחמישה מסמכים
 ב. (4) אם היית צריך לבחור מנוע שישמש כמנוע חיפוש ל web, ומנוע לחיפוש במאגר של פסקי דין, איזה מהמנועים היית בוחר לכל אחת מהמשימות- הסבר מדוע וכיצד זה מתבטא במדדים שחישבת בסעיף א.
 ג. (3) הורצה שאילתא נוספת, המאגר כולל עבורה 10 מסמכים נכונים. להלן תוצאות השאילתא:
 מנוע א': 1, 2, 3, 5, 7
 מנוע ב': 1, 3, 5, 8, 10
 האם ביצועי המנועים עקביים? אם לא, מה יכולה להיות הסיבה וכיצד צריך לבדוק את המנועים כדי להגיע להערכה נכונה שלהם?
 ד. (10) צייר גרף precision recall על פי 11 נקודות recall סטנדרטיות למנוע ב' על פי שתי השאילתות. (ציין את ערכי הנקודות)
 ה. (4) (הסבר את היתרונות והחסרונות של שיטת ה pooling לקביעת relevance judgments במאגרי TREC.
 2. (20) לפניך גרף המייצג רשת Web, בו הצמתים מייצגים דפים והקשתות קישורים בין הדפים.



שאלה 4 (25 נקודות):

לפניך גרף המייצג רשת Web, בו הצמתים מייצגים דפים, והקשתות מייצגות קישורים בין הדפים.



1. (5 נקודות)

איזה דף יקבל את ציון ה-Authority המירבי, ואיזה דף יקבל את ציון ה-Hub המירבי?

כמו בשאלה הראשונה, גם כאןדרושה רק הבנה נכונה של המושגים ואין צורך בחישובים:

המועמדים המובילים לציון ה-Authority המירבי הם: C ו-H כי להם יש את המספר הקשתות הנכנסות הגדול ביותר (3).

שתיים מהקשתות הנכנסות ל-C ול-H זהות: מ-E ומ-D.

בנוסף ל-C, יש A שמצביע עליו ול-H יש את J.

A "תורם" פחות לציון של C מאשר I תורם לציון של H מכיון של-A ציון HUB נמוך משל I. (A מצביע בנוסף רק ל-B שהוא Authority מאוד נמוך לעומת זאת I מצביע ל-F שהוא Authority יחסית טוב).

על כן H הוא ה-Authority עם ציון מרבי.

אותו הגיון לחלוטין יתן את D כ-Hub עם ציון מרבי.

	מיון 2	מיון 1
precision	$\frac{8}{20}$	$\frac{5}{20}$
recall	$\frac{8}{8}$	$\frac{5}{8}$
r-precision	$\frac{2}{8}$	$\frac{4}{8}$
precision 5 doc	$\frac{0}{5}$	$\frac{3}{5}$

7) מיון 5 web - מיון 1, מיון 2, מיון 3
 חישובי, מיון 1, מיון 2, מיון 3
 מיון 1, מיון 2, מיון 3
 מיון 1, מיון 2, מיון 3

8) מיון 1, מיון 2, מיון 3
 מיון 1, מיון 2, מיון 3
 מיון 1, מיון 2, מיון 3
 מיון 1, מיון 2, מיון 3

9) מיון 1, מיון 2, מיון 3

10) מיון 1, מיון 2, מיון 3
 מיון 1, מיון 2, מיון 3
 מיון 1, מיון 2, מיון 3
 מיון 1, מיון 2, מיון 3

2

3) אם נחבר cluster הריבוי עם complete link נשגב גם אלוטריה
 אחר כך נבדוק clusters האם ניתן להפוך, כאשר ה cluster
 המיוחס יהיה באם ניתן להפוך (אם) המיוחס ה cluster
 או cluster אחר.

1-3-2

	T_1, T_2	T_3	T_4	T_5
T_1, T_2	0			
T_3	0	11		
T_4			0	
T_5	3	5	2	0

T_1 ו- T_2 באם ניתן להפוך
 אכן נשגב

	(T_1, T_2)	(T_3, T_4)	T_5
(T_1, T_2)	0		
(T_3, T_4)	0	0	
T_5	3	2	0

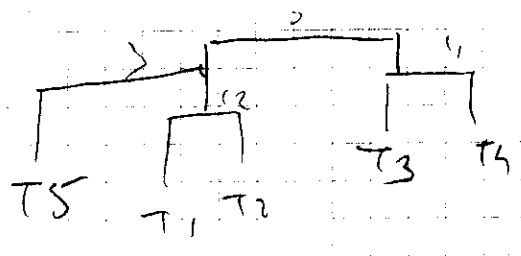
2-3-1

T_3 ו- T_4 באם ניתן להפוך
 אכן נשגב

	(T_1, T_2, T_3)	(T_3, T_4)	T_5
(T_1, T_2, T_3)	0		
(T_3, T_4)	0	0	

3-2-1

T_5 ו- (T_1, T_2) באם ניתן להפוך
 אכן נשגב



$$D_1: \left(\frac{1}{3} \times \frac{1}{3}, \frac{2}{3} \times \frac{1}{5} \right) = \left(\frac{1}{9}, \frac{2}{15} \right)$$

(4)

$$D_2: \left(\frac{3}{4} \cdot \frac{1}{3}, \frac{1}{4} \cdot \frac{1}{9} \right) = \left(\frac{1}{4}, \frac{1}{36} \right)$$

$$D_3: \left(0, 1 \cdot \frac{1}{4} \right) = \left(0, \frac{1}{4} \right)$$

$$D_4: \left(0, \frac{1}{4} \right) = \left(0, \frac{1}{4} \right)$$

$$D_5: \left(1 \cdot \frac{1}{3}, 0 \right) = \left(\frac{1}{3}, 0 \right)$$

$$\text{sim}(Q, D_1) = \left(\frac{1}{9}, \frac{2}{15} \right) (1, 1) = \frac{1}{9} + \frac{2}{15} = 0.009138$$

$$\text{sim}(Q, D_2) = \left(\frac{1}{4}, \frac{1}{36} \right) (1, 1) = \frac{1}{4} + \frac{1}{36} = 1.56$$

$$\text{sim}(Q, D_3) = \left(0, \frac{1}{4} \right) (1, 1) = \frac{1}{4} = 1.25$$

$$\text{sim}(Q, D_4) = \frac{1}{4} = 1.25$$

$$\text{sim}(Q, D_5) = \frac{1}{3} = 1.66$$

المجموع 1.5 - D_3, D_5 المجموع D_1, D_2 (3)

$$\alpha = 1, \beta = 0.5, \gamma = 0$$

$$Q = (1, 1) + \frac{1}{2} \cdot \frac{1}{2} [(0.5, 1) + (1, 1)] =$$

$$Q = (1, 1) + \frac{1}{4} (1.5, 2) = (1, 1) + (0.375, 0.5)$$

$$= (1.375, 1.5)$$