



-8.3

-8.6

-5.2

הוראות לנבחן בצידו השני של הדף

אין לכתוב מעבר לשוליים משני צידי הדף

מס' כיתה 230 בנין 90
מס' נבחן 023286

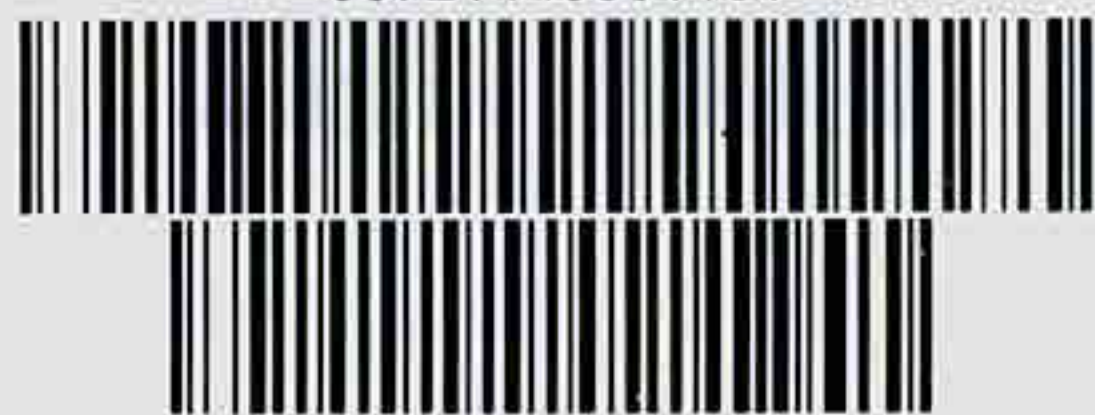
מדבקה

29.03.2009 972373 1 4/4

הנדסת מערכות מידע

אחזור מידע וספריות דיגיטליות

03721440601101



מחלקה הנדסת מערכות מידע שנה ב'
תאריך בחינה 29.3.09
מקצוע בחינה אקטור מידע

לשימוש המרצה הבודק

יחידות | עשרות | מאות

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9

ציון הבחינה 79
שם _____
חתימה _____
תאריך 29.3.09



הוראות לנבחן

1. בהגיעך למקומך יש להניח את כרטיס הנבחן ותעודה מזהה על שולחןך.
2. אסור להביא למקום הבחינה תיקים, ספרים, מחברות, טלפון נייד או רשימות פרט למותר על פי שאלון הבחינה.
3. עזב תלמיד את האולם אחרי חלוקת השאלונים, דינו כדין "נבחן" בבחינה.
4. אסור לנבחן לשוחח בזמן הבחינה, או לעזוב את מקומו ללא נטילת רשות.
5. יש לכתוב בעט, לא בעיפרון. בכתב ברור ונקי, על עמוד אחד של כל דף. אין לכתוב בשוליים ואין להשתמש בטיפקס. טיוטה תכתב בצד ימין של המחברת ויש למחוק אותה בהעברת קו. אסור לתלוש דפים מן המחברת.
6. עבר הנבחן על תקנות הבחינות, יורשה לסיים הבחינה ועניינו יועבר לועדת המשמעת של האוניברסיטה.
7. משך הבחינה מצוין בראש השאלון. עם הודעת המשגיח/ה כי תם הזמן, על הנבחן להפסיק את הכתיבה ולהניח את השאלון על יד המחברת.
8. עם החזרת המחברת לידי המשגיח/ה, יינתן לך אישור על מסירתה. אישור נוסף יימסר תמורת תשלום.
9. אסור להוציא מחברת מחדר הבחינה ויש להחזירה לידי המשגיח/ה לפני עזיבת המקום. אי החזרת מחברת תחשב כעבירה על המשמעת.
10. אין להעתיק חומר כלשהו משאלון הבחינה, אלא אם ניתן לכך היתר מפורש.
11. אין לרשום בשוליים, השוליים נחתכים לצורך סריקה.

בהצלחה.

בס"ד

מבחן בקורס איחזור מידע וספריות דיגיטליות – ד"ר ברכה שפירא, ליהי נעמני
תשס"ט סמסטר א' מועד א', 13:30 29.03.09

חומר עזר מודפס מותר - לא מחשב נייד
משך המבחן: שעה וחצי
יש לענות על כל השאלות.

1. 15% כחלק מתהליך הכנת מאגר להערכת מנועי איחזור, נתנו לשני שופטים להעריך רלוונטיות של מסמכים עבור שאלתא מסויימת. להלן תוצאות הערכתם ל- 4 מסמכים:

מסמך 4	מסמך 3	מסמך 2	מסמך 1	
לא רלוונטי	רלוונטי	רלוונטי	רלוונטי	שופט 1
רלוונטי	רלוונטי	לא רלוונטי	רלוונטי	שופט 2

על סמך נתונים אלה, בחר את המשפט הנכון משלושת המשפטים הבאים (בסס את על חישוב):

- א. בין השופטים יש הסכמה טובה
ב. ההסכמה בין השופטים אינה טובה
ג. אי אפשר להסיק על ההתאמה בין השופטים

2. 20% למנוע חיפוש מסויים יש אינפורמציה על מספר הכניסות ביום של כל דף שמופיע באינדקס שלו.

- א. 10% הצע נוסחת דירוג המתחשבת גם באינפורמציה זו. הנח שנוסחת דירוג הנוכחית של המנוע כוללת כבר מרכיב IR ומרכיב של מבנה הרשת (כגון: pagerank). הנוסחא החדשה צריכה לשלב את כל המרכיבים. הסבר את הנוסחא שהצעת.
ב. 10% כיצד ניתן להעריך האם הנוסחא החדשה משפרת את תוצאות החיפוש? (הצע שיטת הערכה ישימה).

3. 10% אחת השיטות להערכת גודל יחסי של מנועי חיפוש ב web היא לדגום דפים המאונדקסים במנוע אחד ולבדוק האם הם נמצאים באינדקס של המנוע השני. ציין שתי סיבות להטיה אפשרית של הבדיקה הזו.

100

הוא יחיד ארבע ספרים 10^4 ו- 10^5 הם חזקות של 10. כלומר, $10^4 = 10 \cdot 10 \cdot 10 \cdot 10$ ו- $10^5 = 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10$.

הוא יחיד ארבע ספרים 10^4 ו- 10^5 הם חזקות של 10. כלומר, $10^4 = 10 \cdot 10 \cdot 10 \cdot 10$ ו- $10^5 = 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10$.

- הוא יחיד ארבע ספרים 10^4 ו- 10^5 הם חזקות של 10. כלומר, $10^4 = 10 \cdot 10 \cdot 10 \cdot 10$ ו- $10^5 = 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10$.

$$\frac{10^5}{10^4} = \frac{10 \cdot 10 \cdot 10 \cdot 10 \cdot 10}{10 \cdot 10 \cdot 10 \cdot 10} = 10$$

הוא יחיד ארבע ספרים 10^4 ו- 10^5 הם חזקות של 10. כלומר, $10^4 = 10 \cdot 10 \cdot 10 \cdot 10$ ו- $10^5 = 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10$.

- הוא יחיד ארבע ספרים 10^4 ו- 10^5 הם חזקות של 10. כלומר, $10^4 = 10 \cdot 10 \cdot 10 \cdot 10$ ו- $10^5 = 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10$.

הוא יחיד ארבע ספרים 10^4 ו- 10^5 הם חזקות של 10. כלומר, $10^4 = 10 \cdot 10 \cdot 10 \cdot 10$ ו- $10^5 = 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10$.

- הוא יחיד ארבע ספרים 10^4 ו- 10^5 הם חזקות של 10. כלומר, $10^4 = 10 \cdot 10 \cdot 10 \cdot 10$ ו- $10^5 = 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10$.

4. 20% שני מנועי חיפוש: e_1, e_2 הורצו על שתי שאילתות: Q1, Q2. ידוע שלשאילתא Q1 10 מסמכים רלוונטים במאגר, לא ידוע מספר המסמכים הרלוונטים במאגר לשאילתא Q2. להלן תוצאות ההרצה של השאילתא לשני המנועים (+ מסמן מסמך רלוונטי שהוחזר, - מסמן מסמך לא רלוונטי):

מנוע E2		מנוע E1	
Q2	Q1	Q2	Q1
+	-	+	+
-	+	+	+
-	-	+	+
+	-	+	+
-	+	-	+
+	+	+	-
+	+	-	-
+	+	-	-
+	+	-	-
-	+	-	-

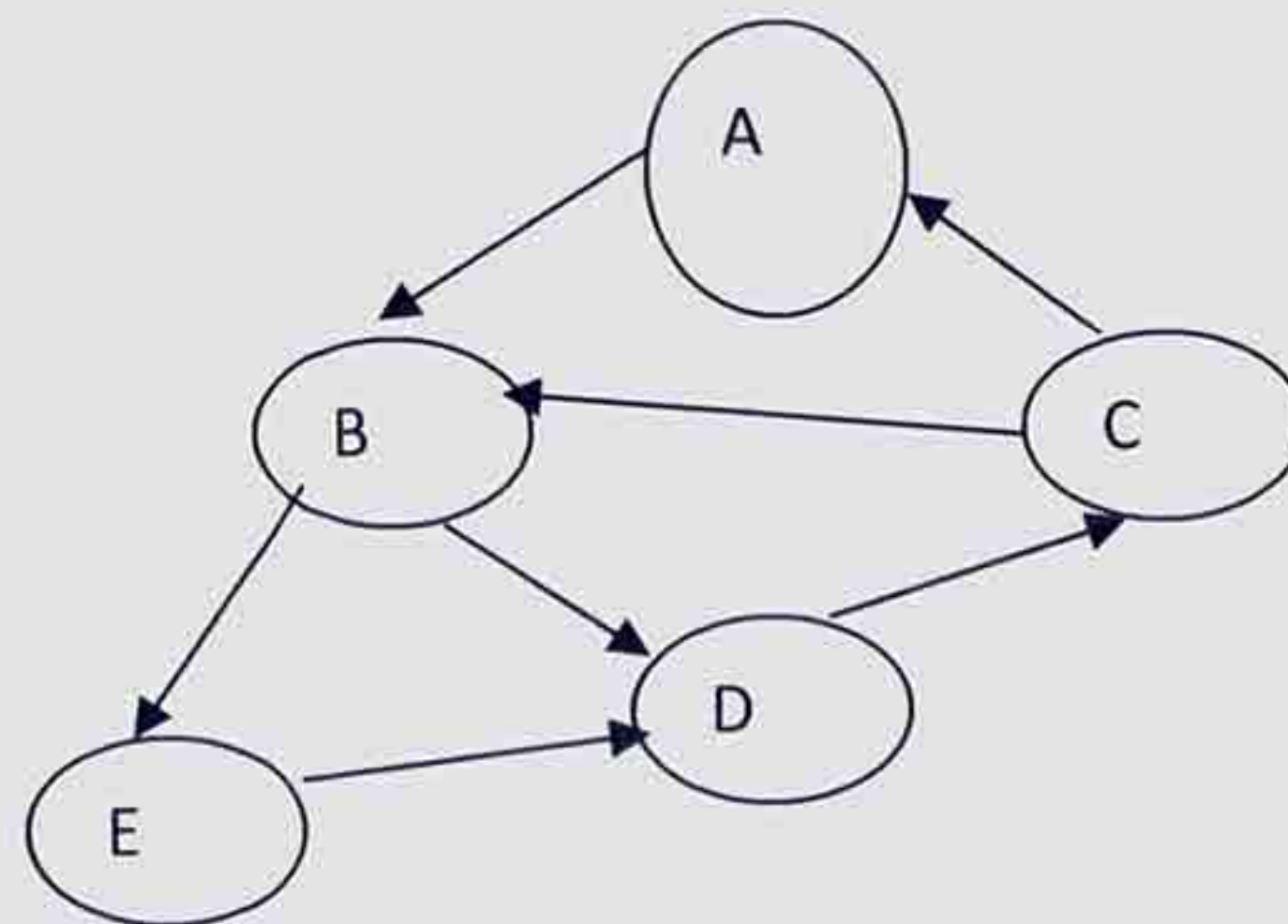
- א. 3% האם ניתן להסיק על מספר המסמכים הרלוונטים הנמצאים במאגר לשאילתא Q2 מתוך התוצאות לעיל?
- ב. חשב, או הסבר למה לא אפשר לחשב את הערכים הבאים:
- a. 3% mean average precision על פני שתי השאילתות למנוע E1.
- b. 3% mean average precision לשאילתא Q1 למנוע E2.
- c. 5% E-measure לשאילתא Q1 למנוע E1 כאשר ניתן משקל כפול ל precision לעומת ה recall.
- ג. 6% אפיין את התנגדות כל אחד מהמנועים, (בהנחה ששתי השאילתות מייצגות את התנהגות הכלליה).

5. 20% בניח מאגר הכולל רק את המסמכים : d1,d2,d3 להלן נתונים על תדירות terms במסמכים:

Terms	D1	D2	D3
Information	35	0	200
Retrieval	0	100	0
course	25	40	0

- א. 6% ייצג את המסמכים באמצעות ווקטור משקולות על פי נוסחת $tf \cdot idf$ יש לנרמל את tf ל term השכיח ביותר. הראה את כל החישוב שלך.
- ב. 6% אילו שני מסמכים קרובים ביותר ביניהם? (השתמש בנוסחת קוסינוס).
- ג. 8% מדוע כדאי לחשב דמיון בין מסמכים במאגר? (תן דוגמא לשימוש אפשרי).

6. 15% נתונה הרשת הבאה:



- א. 5% מהו הצומת/הצמתים בעלי ה pagerank הנמוך ביותר (אין צורך לחשב, אלא רק להעריך). הסבר את הערכתך.
- ב. 5% הסבר איך ישפיע ביטול הקישור בין C ל-A על ערכי ה Pagerank ברשת? (אין צורך לחשב, אלא רק להעריך)
- ג. 5% הסבר איך ישפיע ביטול הקישור בין C ל B (בנוסף על הביטול בסעיף ב') על ערכי ה pagerank ברשת? (אין צורך לחשב, אלא רק להעריך)
- ד.

בהצלחה

ברכה וליהי

השאלה היא: האם יש קשר בין המשתנים? אם כן, מהו סוג הקשר?

משתנה	א	ב
1	1	2
2	2	3
3	3	4
4	4	5

1. האם יש קשר בין המשתנים? אם כן, מהו סוג הקשר?
2. האם יש קשר בין המשתנים? אם כן, מהו סוג הקשר?
3. האם יש קשר בין המשתנים? אם כן, מהו סוג הקשר?
4. האם יש קשר בין המשתנים? אם כן, מהו סוג הקשר?
5. האם יש קשר בין המשתנים? אם כן, מהו סוג הקשר?

השאלה היא: האם יש קשר בין המשתנים?

1. האם יש קשר בין המשתנים? אם כן, מהו סוג הקשר?
2. האם יש קשר בין המשתנים? אם כן, מהו סוג הקשר?
3. האם יש קשר בין המשתנים? אם כן, מהו סוג הקשר?
4. האם יש קשר בין המשתנים? אם כן, מהו סוג הקשר?
5. האם יש קשר בין המשתנים? אם כן, מהו סוג הקשר?

השאלה היא:

השאלה היא:

לחלק 1:

נסתף מן קאפא :

$$\text{Kappa} = \frac{P(A) - P(E)}{1 - P(E)}$$

$$P(A) = \frac{2}{4} = 0.5$$

~~P(A)~~

$$P(\text{nonrelevant}) = \frac{2}{8} = \frac{1}{4}$$

$$P(\text{relevant}) = \frac{6}{8} = \frac{3}{4}$$

$$P(E) = \left(\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2 = \frac{1}{16} + \frac{9}{16} = \frac{10}{16} = \frac{5}{8} = 0.625$$

$$\text{Kappa} = \frac{0.5 - 0.625}{1 - 0.625} = \frac{-0.125}{0.375} = -\frac{1}{3}$$

לפי נוסחה - הסתמה
לא סופה

פירוט: אין הסתמה סופה בין השניים (ב').

(ההסתמה בין השניים היא מרובה).

צוויכער סקטור אין - גוילריון בנאג

הם מנהלתי את מדיניות החוץ, נכח ניסוי בו השתתף
 יחדיו אל ~~המנהל~~ המנהל ~~המנהל~~ המנהל, וניסוי אחר
 שקיווה הרכיב של השתתפות.

את השתתפות ניסוי נחש לא רקובות צוות בקולנו, ודוחות מחמית אופי
 השתתפותם הם מנהל למודע את הפרמטר הנמדד. - אלף אלף תוצאה
 מתאליה יק לשון כזה מל מנהל מנהל!
 קבוצה א' - גשון קבוצה בקרה, קבוצה מסמך המדענים לפי הדיווח
 במהלך של מנהל החיפוש (IR ו-PR בלבד).
 קבוצה ב' - קבוצה מסמך המדענים לפי הדיווח שהקבוצה מנהלית
 מוצגת בסעיף א'.

נחש את הניסוי קודם מן אחרת (אחרת צוות קודם של הקבוצה
 וקודם אותם מאחרת) $\frac{M}{N}$ $\frac{M}{N}$

בניסוי יבדקו קודם מן מספר את מנהל הניסוי אליו יחסית לדיווח, נבדוק האם
 מנהל הניסוי אליו מנהל לדיווח (האם למנהל) והמנהל נכנס ואלו המנהל

המנהל במנהל (המנהל) X (מנהל) $\frac{M}{N}$ $\frac{M}{N}$ וזאת במנהל
 כמו כן, נראה את הדיווח של המנהל של precision, recall (במידה
 ובמנהל סגור בלבד) ובמנהל של ידי שקיווה רצון השתתפות (במנהל) אליו
 מספר או האיות קודם בלבד הניסוי.

אלף קבוצה א' אל המנהל למנהל סעיף ב'

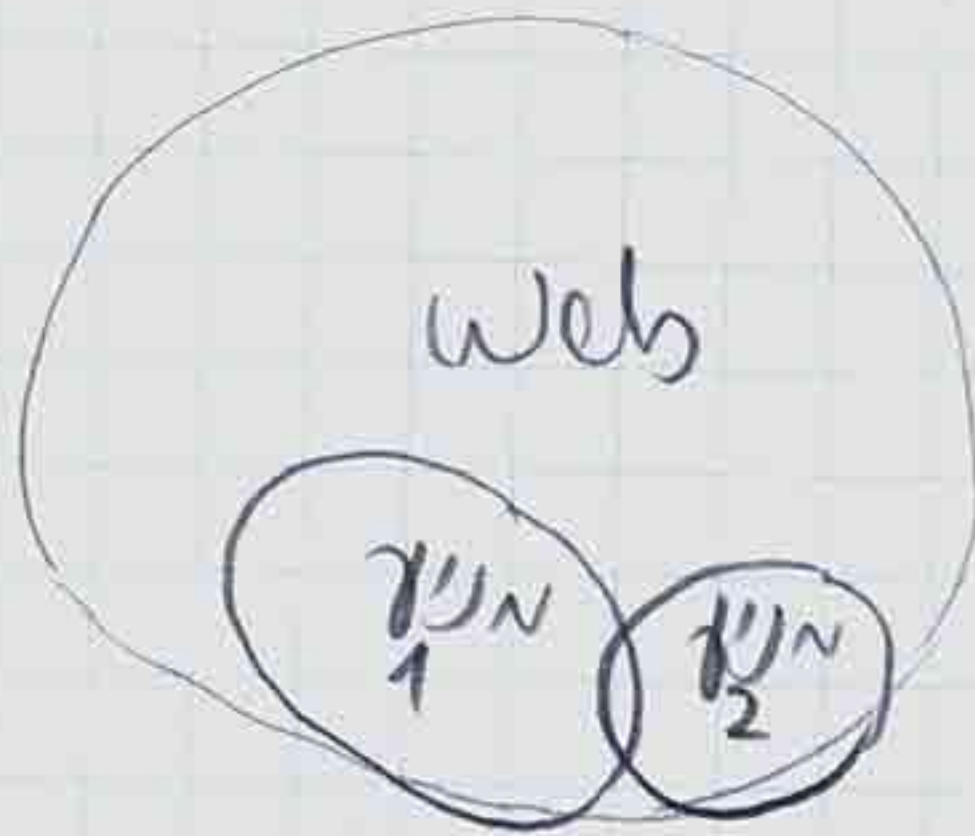
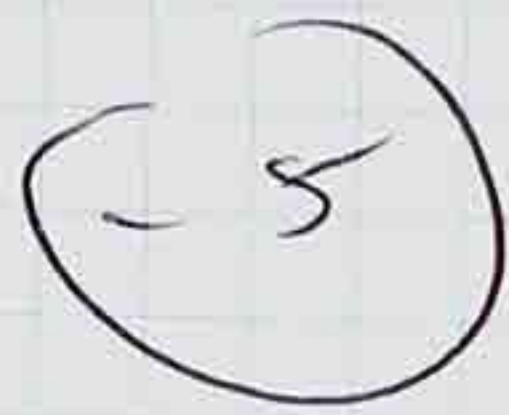
* ניתן לתת כ' בלבד קבוצה יחדיו מספר השתתפות (אם י' ואלו השתתפות רצון) $\frac{M}{N}$
 למנהל השתתפות.

$\frac{M}{N}$ $\frac{M}{N}$ $\frac{M}{N}$

למוד 3

~~למוד 3~~

חומר בין האינטרנט של מנוע שני המכשיר חיפוש ב-web מזהה חסות
 אם ברצוננו להאמת זאת הדבר של האינטרנט בין המנוע השני.
 הסיבה לכך היא שמנוע שני ב-web מאפשר מנוע ב-web. זאת
 'רע' ~~למוד~~ שני חומר מאוזן בין מנוע מזהה ומסומן השורה. ואין
 מאפשר אים שני אחרות מנוע אתר.



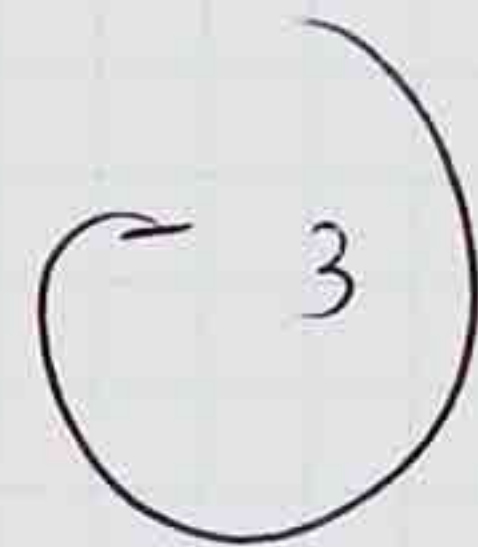
לשם הקדמה לשני
 המנועים זהו קומה

אם A, לא בין המנוע B, זה להבין.

אם נוסף לעומת זאת המנוע המנוע של אחר המנוע.
 יש מנוע אחר המנוע online ואילו המנוע של המנוע מאפשר
 shadowing (כל אחר המנוע, זה המנוע, מנוע אחר המנוע - זה המנוע
 המנוע, ואם המנוע מנוע switch).

מנוע זה, יש מנוע בין המנוע המנוע online למנוע המנוע
 (זה המנוע), ואם זה המנוע זה המנוע המנוע המנוע
 המנוע, המנוע המנוע לא המנוע המנוע.

למנוע המנוע המנוע המנוע



מנוע המנוע המנוע המנוע
 המנוע המנוע המנוע המנוע
 המנוע המנוע המנוע המנוע
 המנוע המנוע המנוע המנוע

שאלות ותשובות

אונברסיטת סמיטת - גורליוז בנגב

10 - 9, 10 - 9, 10 - 9

	2 JUN				1 JUN			
	Q ₂		Q ₁		Q ₂		Q ₁	
	PR	recall	PR	recall	PR	recall	PR	recall
1	1		0	0	1		1	0.1
2	0.5		0.5	0.1	1		1	0.2
3	$\frac{1}{3}$		0.333	0.1	1		1	0.3
4	0.5		0.25	0.1	1		1	0.4
5	0.4		0.4	0.2	0.8		1	0.5
6	0.5		0.5	0.3	0.833		0.833	0.5
7	0.571		0.571	0.4	0.714		0.714	0.5
8	0.625		0.625	0.5	0.625		0.625	0.5
9	0.666		0.666	0.6	0.555		0.555	0.5
10	0.4		0.4	0.7	0.5		0.5	0.5

$\therefore E2$ jump of control MAP sens zu (b)

$$\text{MAP}(q, r) = \frac{0.5 + 0.4 + 0.5 + 0.571 + 0.625 + 0.666 + 0.7}{7} = 0.566$$



(c) E-measure definition of μ and σ is:

$$PR(e_1) = 0.5$$

$$\text{Recall}(e_1) = 0.5$$

precision, recall \rightarrow \rightarrow \rightarrow precision-recall graph

$$\beta^2 = 2$$

$$E = \frac{(1+\beta^2) \cdot P \cdot R}{\beta^2 P + R} = \frac{(1+2) \cdot 0.5 \cdot 0.5}{2 \cdot 0.5 + 0.5} = \frac{3 \cdot 0.5 \cdot 0.5}{1 + 0.5} = \frac{0.75}{1.5} = 0.5$$



- 2

מידת הצפייה של Precision היא מידת הצפייה של המידות
 (Precision) מידת הצפייה של המידות.
 מידת הצפייה של המידות (Precision) מידת הצפייה של המידות.
 מידת הצפייה של המידות (Precision) מידת הצפייה של המידות.
 מידת הצפייה של המידות (Precision) מידת הצפייה של המידות.

מיון עז הכולל מיון לפי Precision ומוקד יותר במיון המסווג, המיון והוא מבוצע על ידי מודל מאונסר.

מודל, אם נוסף מיון של recall - זה יהיה (המיון העדיף).

מיון, מיון מסווג - Recall זה יחיד.

הערה, מניח ש אתר מיועד להיות מיועד לweb, וזה מיועד להיות
מיועד להיות מיועד לweb (הערה) מיועד להיות מיועד לweb.

שאלה 5:

	d1	d2	d3
Information	35	0	200
Retrieval	0	100	0
Course	25	40	0

מספר המופעים של המילה term - המילה קובע:

	d1	d2	d3	df	idf $\leftarrow \log_2 \left(\frac{N}{df} \right)$
Information	1	0	1	2	0.584
Retrieval	0	1	0	1	1.584
Course	0.414	0.4	0	2	0.584

המספרים המוקדמים (w=tf*idf) המוקדמים

	d1	d2	d3
Information	0.584	0	0.584
Retrieval	0	1.584	0
Course	0.417	0.2336	0

הוקדמו המוקדמים tf-idf שהם:

$d_1 = (0.584, 0, 0.417)$

$d_2 = (0, 1.584, 0.2336)$

$d_3 = (0.584, 0, 0)$



$$D_3: \sum w_{ij}^2 = 0.584^2 = 0.341$$

$$\cos \text{Sim}(D_1, D_2) = \frac{0.417 \cdot 0.2336}{\sqrt{0.514 \cdot 2.563}} = \frac{0.0974}{1.1477} = 0.084$$

$$\text{CosSim}(D_1, D_3) = \frac{0.584 \cdot 0.584}{\sqrt{0.514 \cdot 0.314}} = \frac{0.341}{0.4017} = 0.848$$

$$\text{cosSim}(D_2, D_3) = \frac{0}{\sqrt{2.563 \cdot 0.341}} = 0$$

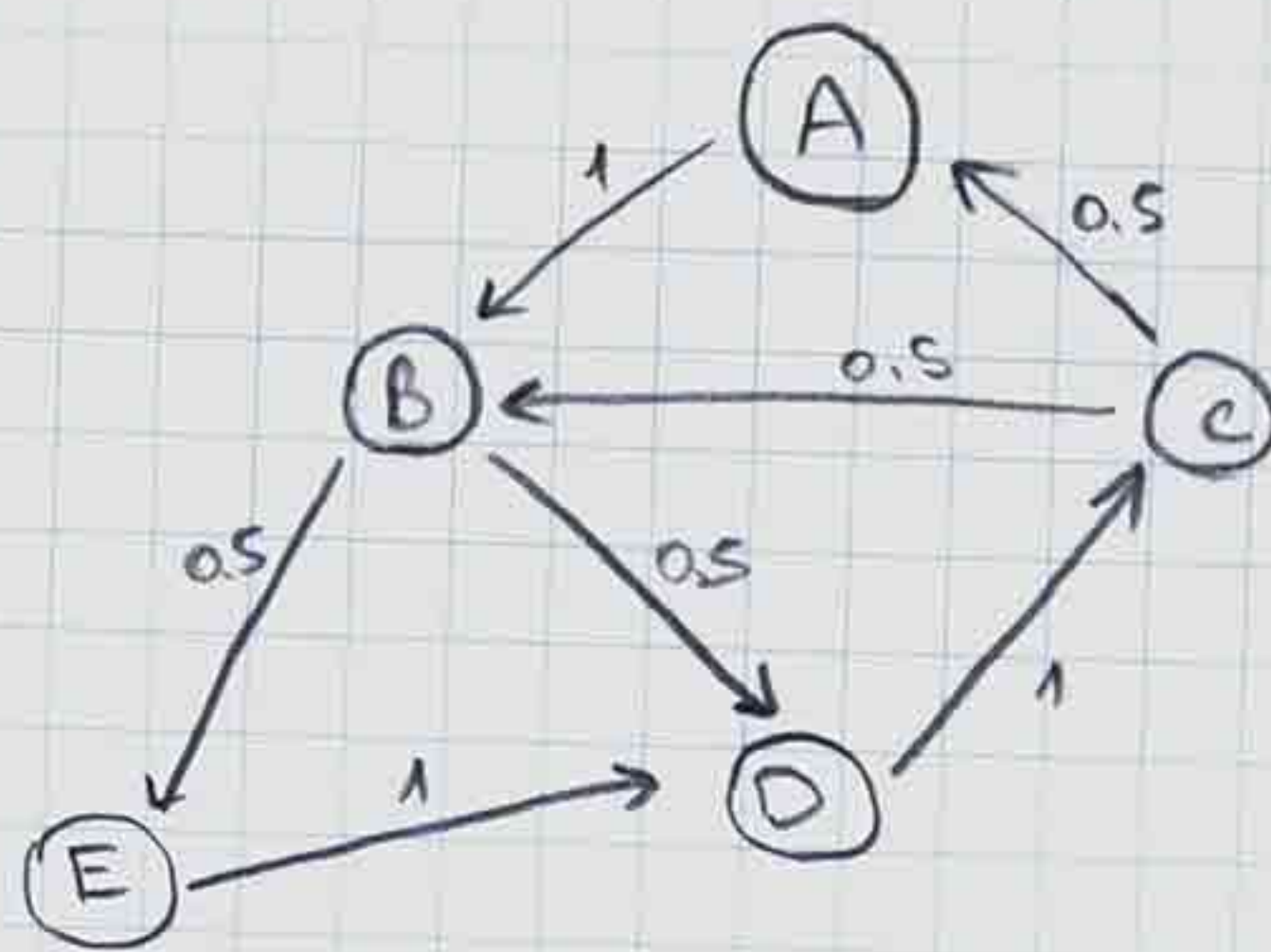
המרחב הקרוס קיורת קיורת d_1, d_3

- 2. Pb^{2+} and Fe^{3+} ions are present in the solution. The solution is acidic. The Pb^{2+} ions are precipitated as PbCl_2 by adding HCl . The Fe^{3+} ions are precipitated as Fe(OH)_3 by adding NaOH . The PbCl_2 is dissolved in HCl and the Fe(OH)_3 is dissolved in NaOH . The Pb^{2+} ions are precipitated as PbSO_4 by adding H_2SO_4 . The Fe^{3+} ions are precipitated as Fe(OH)_3 by adding NaOH . The PbSO_4 is dissolved in HCl and the Fe(OH)_3 is dissolved in NaOH .

רשומה: ייחודיות של פתרון מסווג (גורם מסווג) בין שני מסמכים, ולפיכך המסמך
 מוצגים המיושם רק על מסמכים מסווגים ~~מבוא~~ למסמכים, ומסמך שני צומח לו
 מן ~~ה~~ הפתרון, ומכאן יש את המסמך השני כמובא המיושם, על מנת להבין
 לא מבוא ~~למסמכים~~ למסמכים (רשומה אם מושגים בו מילי נרשם למילי
 (המילי) על מילי נרשם מן הפתרון, ומכאן למילי המילי.

[illegible]

שאלה 6:



-כ

$$PR(A) = 0.5 PR(C)$$

$$PR(B) = PR(A) + 0.5 PR(C)$$

$$PR(C) = PR(D)$$

$$PR(D) = 0.5 PR(B) + PR(E)$$

$$PR(E) = 0.5 PR(B)$$

$$PR(A) = 0.5 PR(C) \Rightarrow PR(A) < PR(C)$$

$$PR(B) = PR(A) + 0.5 PR(C) \Rightarrow PR(A) < PR(B)$$

$$PR(D) = 0.5 PR(B) + PR(E) \Rightarrow PR(E) < PR(D)$$

$$PR(E) = 0.5 PR(B) \Rightarrow PR(E) < PR(B)$$

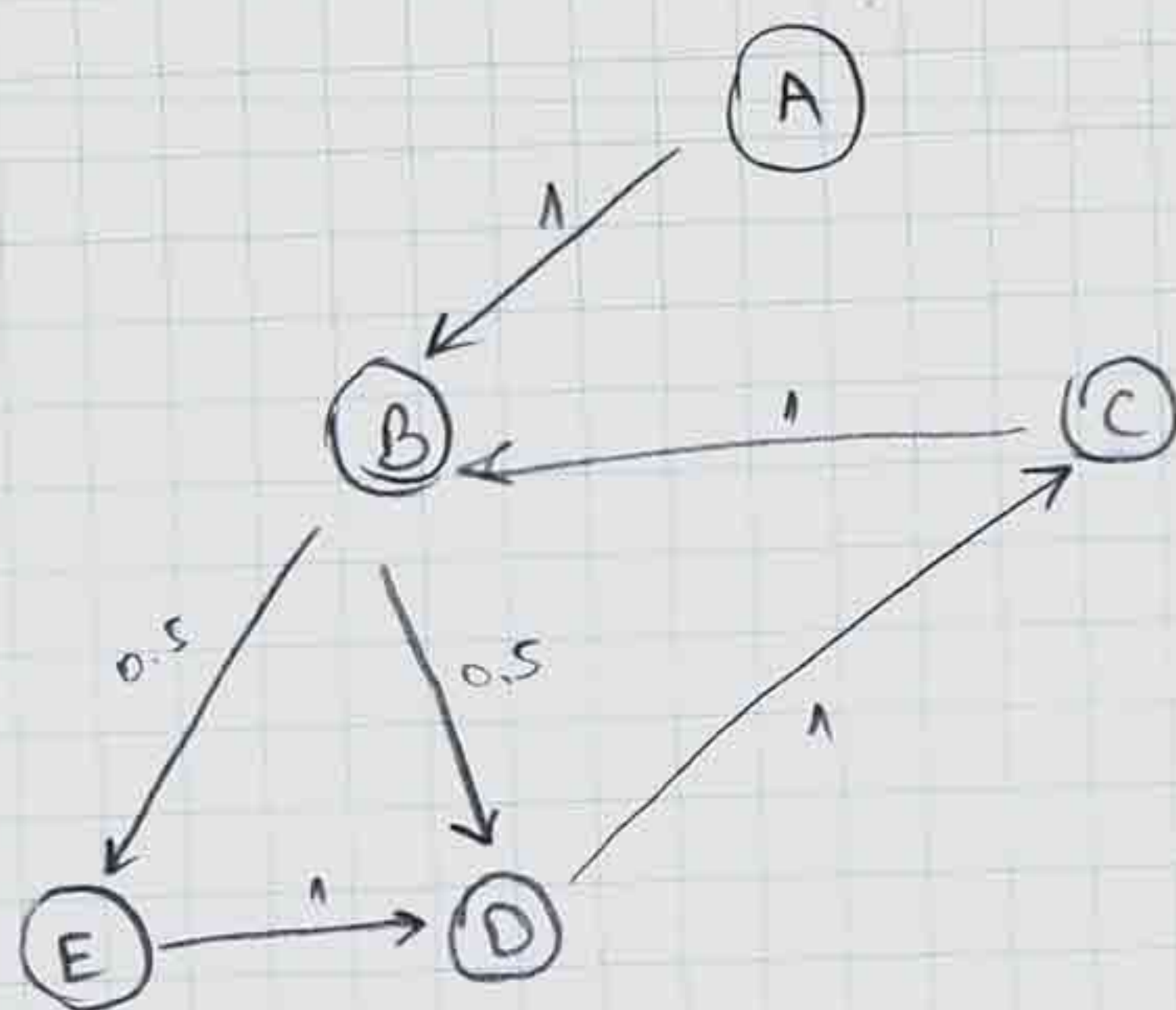
$C > A$
 $B > A$
 $D > A$

$D > E$
 $C > E$
 $B > E$



למחרת: הסתכלו בקוד PR - קיבלו קיולט על צורת A ו-E
 אולם קיולט מראה על A ו-E - נראה שיש פה טעות, אולי

צורכי סביבת בני-גורליות בגב

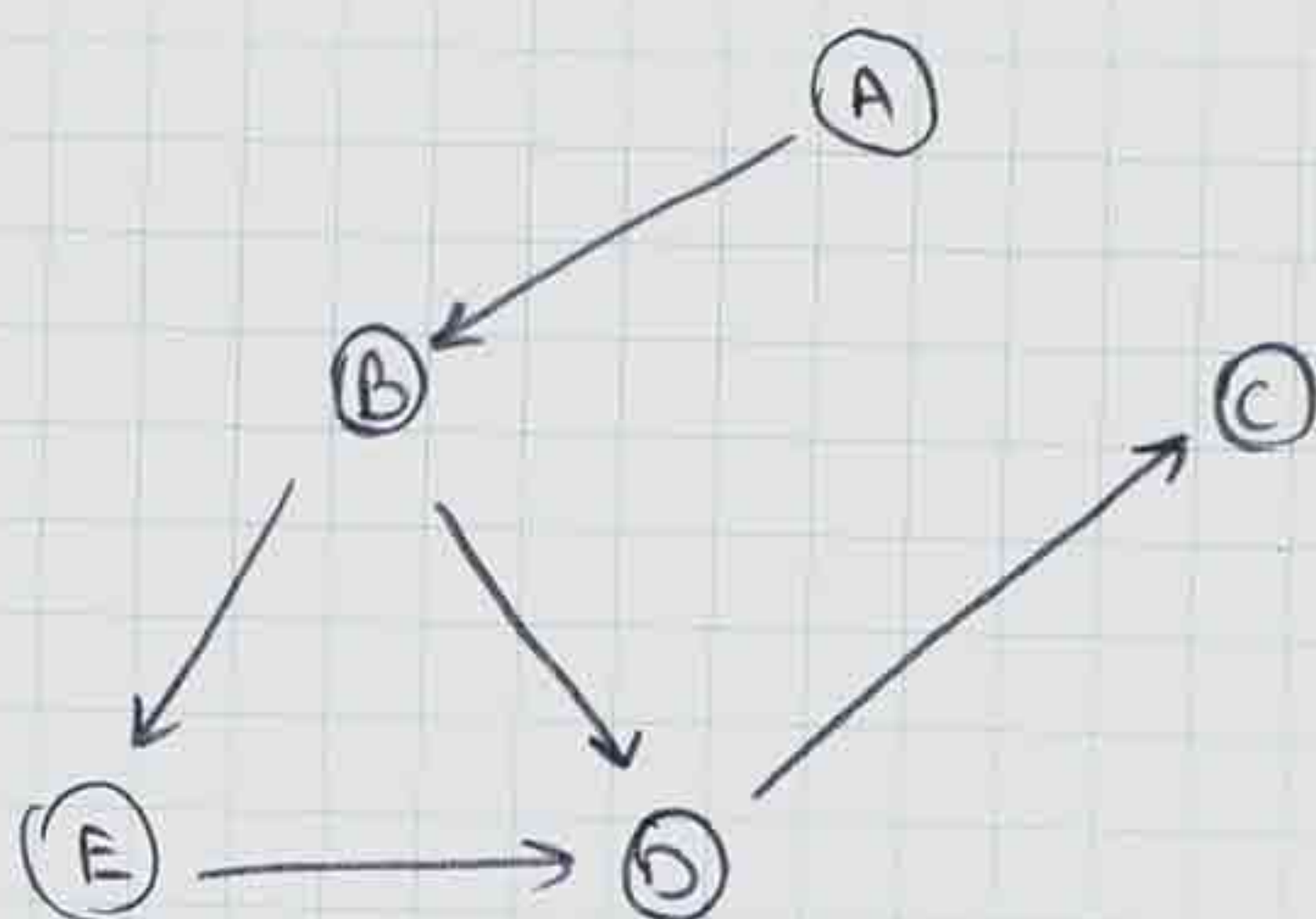


~~ה-PR של A יורד ויורד עד ויגיעה למסלול המינימלי~~
~~לדלג ממנה וזה קטן, לאבר מסלול מינימלי A יורד עד קטן קטן B-1~~

מחיבה והתעורר פילגה ק שיהא שאורה לא קוצהק קטן (1-d), פה
 מאינציה מסוימת, ערנו של A יפה קטן, ופא יור קטן B-1.
 מחיבה והתעורר פילגה ק שיהא שאורה לא קוצהק (1-d) PR (אחרי יחס)
 מה-PR, ערנו של ~~PR~~ קוצהק A יור ויקטן כל קטן באינציה
 בעדמות פילגה מ, A יופקרוהיות הם ה-PR הנחוק סוף קטן.

X לא מספיק נכון (-3)

הירשאה הוא ש A נטא יור
 הקקא (1-d) קטן
 מחלקה קאקאם אה ה pagerank



מיטל הקשר בין B-SC (מקור) & מיטל הקשר בין A-SC
 ימים זמן ל-C יהווה dead end כולל. למטה - הוא יקרה אל PR
 הקרה מילש מכיוון שבזמן בין יקרה, ולא יזן אם צברה

[illegible]

הנהגה

5

3.7.2

האנו מודים לך מאוד על
העזרה והתמיכה שאתה נותנת לנו.
נשמח להמשיך ולקבל את
העזרה שלך.

בטוב,
ד"ר אביחי גורן

מתוך המסמכים האלו על PR -הן PR

(1d)-PR (כמו מסמכים אחרים)

8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100

אונברסיטת סמינר לב זול-גורליון באגב

צוויי סקטען בז-גורליוז בונגב

צאצאב לאוראז-אן-גורדיז בן-ספיטא

צווייבליטער בוך - גורליוז בנגב

שאלות ותשובות - ארבעה זמנים

אונברסיטת סמינר בז-גורליוז בנגב

Directly Observed Therapy

