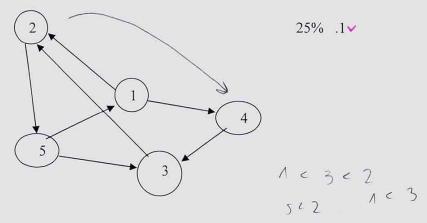
## 111 70 36 1 8/8

בס"ד

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

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



- ו HUB אם נוסיף קישור מצומת 2 לצומת 4, האם וכיצד יושפעו ערכי  $\checkmark$  אם נוסיף קישור מצומת 5 (כלומר , האם הערכים יהיו גבוהים, AUTHORITY (על פי אלגוריתם או נמוכים יותר, או ללא שינוי ומדוע) .
  - ישל הרשת מסעיף א (לאחר (לצורך חישוב pagerank) ב. עב. (5%) הראה את מטריצת המעברים (לצורך חישוב ההוספה של הקישור מסעיף א).
- ער. (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 בגף הנתון ביותר לאחר ההתכנסות. אם כן הערך והסבר את הערכתך, או הסבר מדוע אי אפשר להעריך.
- ה. (5%) האם אלגוריתמים של link-analysis כדוגמת האם אלגוריתמים של האוריות של בים את פופולאריות הדף אצל הגולשים. אם כן, כיצד? ואם לא איך אפשר למדוד פופולאריות של דפים אצל נולשים?
- 2. 8% רשומה בקובץ posting הופכי כוללת נתונים על פי terms. על כל posting כולל את רשימת כל המסמכים שבהם הוא מופיע, ואינפורמציה נוספת (למשל, מיקום בולל את רשימת כל המסמכים). 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 שבהן ה-min term יותר משמעותי יופיעו ברשימה לפני מסמכים שבהם ה term פחות משמעותי.

ציין יתרון אחד לכל אחת מהשיטות.

2 0.312 2 0.6 3 0.6 4 0.4 PM

1

www.bgu4u.co.il

מאגר הסיכומים

באדיבות אגודת הסטודנטים, בן גוריון

## M1 70 36 1 8/8

ne mind web the

. 3 <u>עום מטה-מנועים ממזגים תוצאות של כמה מנועי חיפוש לתוצאה ממוינת אחת שמוצגת</u> למשתמש

א. א. 5% מיזוג תוצאות נחשבת לבעיה שמצריכה אלגוריתם מורכב. הסבר ממה נובע הקושי במיזוג התוצאות.

ב. 5% אם כל מנועי החיפוש שמשתתפים במטה-מנוע ישתמשו באותו אלגוריתם דרוג. האם הקושי שתיארת בסעיף א יימנע? הסבר מדוע כן או לא.

4. "4 הנח שאתה בונה מנוע המבוסס על המודל הווקטורי. בנוסף הדרישה היא שפונקציית הדרוג תתבסס על שלושה סוגי מידע: 1) הטקסט שבגוף המסמך, 2) כותרת המסמך, 3) טקסט של קישורים (anchor text).

א. 8% הסבר כיצד אפשר לשלב את המודל הווקטורו כך ששלושת סוגי המידע יבואו לידי ביטוי בחישוב הדרוג של המסמך וחישוב הדמיון יהיה על סמך התאמה בין ווקטור השאילתא לווקטור המסמך. סוגי המידע יכולים להשפיע באופן שונה על הדירוג, (יתכן שסוג מידע אחד יהיה משמעותי יותר מאחר).

ב. 10% על פי תשובתך ב-א, תאר פונקציית דירוג ( תיאור בנוסחה) על פי הכללים ב. אפשר לתאר פונקציה נפרדת לכל אחד מהכללים) :

אם מילת חיפוש מופיעה בכותרת, לא מתחשבים במספר ההופעות שלה בגוף הטקסט או בטקסט של קישורים.

אם מילת חיפוש לא מופיעה בכותרת, אז מתחשבים בתדירות הופעתה במסמך ובטקסט של קישורים, כאשר טקסט של קישורים משפיע פחות מטקסט בגוף המסמך.

R = 30 P = 10

- .5 שהחזיר. 10 תוצאות שהחזיר. 01 תוצאות נכונות לשאילתא Q1 מתוך 20 תוצאות שהחזיר. סה"כ יש במאגר 30 תוצאות מתאימות לשאילתא. ובכלל במאגר יש 1000 מסמכים. הגדילו את המאגר ל10,000 מסמכים. למאגר לא התווספו עוד תוצאות רלוונטיות לשאילתא Q1 (כלומר, נשארו 30 תוצאות מתאימות). הריצו מנוע אחר E2 על המאגר החדש והוא החזיר בדיוק את אותם 20 מסמכים שהחזיר המנוע הראשון על המאגר הקטן יותר: סמן "נכון" או "לא נכון" על כל אחד מהמשפטים הבאים:
- על נכון אוים אילתא Q1 אילתא עני המנועים של recall אוים אוים אילתא ער אילתא אילתא אילתא איכון איכון איכון איכות באלף איכות המנועים. איכות המנועים. איכות המנועים. איכות המנועים. איכות המנועים. איכות המנועים. איכות המנועים.

נכון/ לא נכון. ענורמל למספר המסמכים במאגר יבטא את ההבדל בין איכות Precision 2% ... ער.

יחזיר Q1 אם יופעל על המאגר הקטן (1000) עם שאילתא Q1 יחזיר , אם יופעל על המאגר הקטן (1000) עם שאילתא יחזיר (15 תוצאות רלוונטיות. נכון לא נכון

6. משתמש במערכת סינון תוכנית קיבל כהמלצה 3 מסמכים: d1,d2,d3, המשתמש כתגובה טען שמסמך d1 לא רלוונטי, ואילי d2 ו-d3 רלוונטים מאוד. המערכת מעדכנת את פרופיל המשתמש בהתאמה לתגובותיו . אפשר להסיק ש: (בחר תשובה נכונה אחת).

א. D2 ו-d3 דומים ביניהם .

d3 מאשר ל d2 המסמכים הבאים שיחזרו למשתמש יהיו דומים יותר ל d2 מאשר ל פני
 ג הפרופיל של המשתמש אחרי העדכון יהיה דומה יותר ל d2 ו-d3 מאשר לפני

.d3 ו-d2 מאשר ל d1 – הפרופיל של המשתמש לפני העדכון דומה יותר ל

2

· 10.

gold stu

6.6n. (110

, No. 7 776C

ישון יון

30 JANOUL

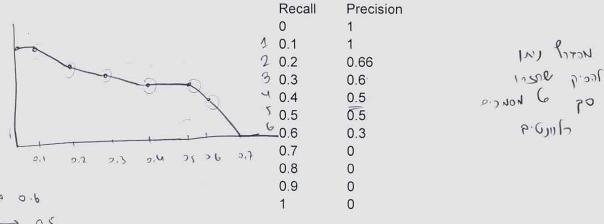
שונים בניווני שווש בניווני

XI EIII -

M 7036 1 8/8

2 0 66

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



MOTO UNI

P. G1110

8 → 0.5  $\frac{90}{200} \rightarrow \frac{6}{20} = 03$ 

P=0.5 P=0.6 ,מסמכים ו -20 מסמכים אחר שהמערכת לאחר שהמערכת פיסוג אחר לאחר פיסוג אחר לאחר היה היה היה אחזירה לאחר שהמערכת לאחר שהמערכת מסמכים, אחר שהמערכת מסמכים, אחר שהמערכת היה אוד של היה אחר שהמערכת היה אוד של היה אחר שהמערכת היה אוד של היה שהמערכת החזירה 5 מסמכים רלוונטיים. precision לאחר שהמערכת

-8√ מוצרים שרכשו (1-5). לצורך ניבוי מוצר 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-

לצורך ניבוי ההתאמה של מוצר 5 למשתמש 3, השכנים שיילקחו בחשבון הם (רק תשובה אחת נכונה):

	7				
	משתמש 1	משתמש 2	משתמש 3	משתמש 4	משתמש 5
מוצר 1	5	/2	5	7	4
מוצר 2				5	
מוצר 3				5	
4 מוצר	3	5	2		3
מוצר 5		5		5	5

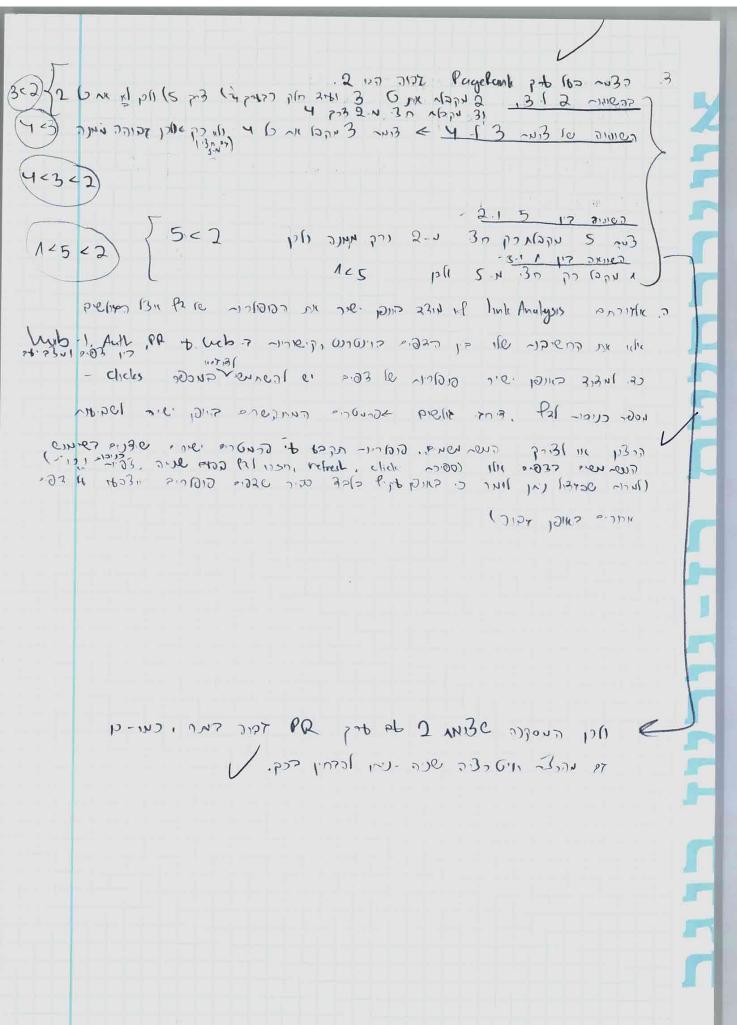
- א. אין שכנים מתאימים
  - ב. משתמשים 2 ו-5
- - משתמש 5
- ה. אף תשובה לא נכונה

-N' 2 WACH 1.34 10/1

GC GCIG 19.1

בהצלחה

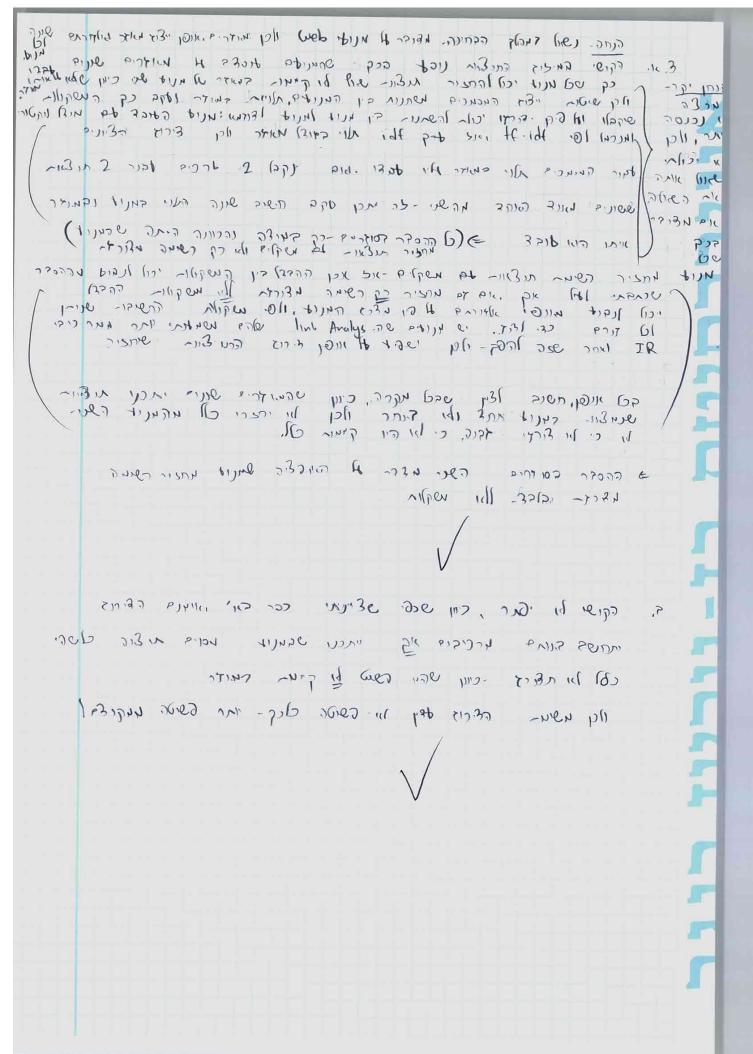
ברכה ואיליה



And the short short short and short for hind short for the short are and the short and

-לוה פון של מונגטימצניה לפור אנילי הלאיניי לה שולאי- וחא

El Megs volge vigo viso viso viso viso



```
אין נכצה לפני בעקציה שתשלב זות שחושה הדורמוים יולכן תחילב נכלב לבפין מבו
          העסקו שו כל דוחם כב לצא אים הן ישביו א הנוכחה ולכן לצורק
          דנית עוכחה קבינון והחוסה ל משהו הפרעטרי לכלל ניסויים א כל פרמטר
   כבי לכיול אנתם דניסוים זכן ניונח לכנה טדורה שפיל. (כיון שציין רשולה שלוי יצוד
 (21, 430 13 4 10 9211
                                            matire los Par Bare 1017
                           Person le moron maion 2 pt
                                            eidin film inik H
                   cq varior ve & copie : Q td 100 00 00
                      ~2010 CIUCA
               anchortext elg
      (Ell) ACMONTAL ES CHO MILL LEE OF (101, MY Eil Eil 01-1
      ופי הצחן בו קובוו אי המנט נים שבמיב רצון מברוווטיו- שורים
                11100 seither the city 120 mil las 19 ticd (1.9
                               שפוז גינל זותצ יכו לכיוה יותר שאלהי אותכ
     2= 0.6
                       प्रमा क्षेत्र मात्र में हार हिल्ला भी कार कार
      B= 0.3
                                        भारती। ने तथदा कर नामिया
      1.0 = 1
                                              ( (m) 16214 0 00 pecie mer 4)
    אותר שעלונו שוהו של דורק נשנה זות העשקו שו הווחשר בעולו המתשרי
( השקבט שבדין העימון - יחושה ום ננסחה שלה כ שוניב (יהה עשה וכם אלחול כ כתבה העימון הוא הנו אלחול
                                   בניקטר הנפן יחב ומשתור שקבאני ואו.
        2/1 (100 1)12 NG POU AD 83"ND 716701 016ND NK Plan -14 A16
     A11 (32) - 1036 00767 - 10714 Ancho - 30 00 767 - 184 DX - Anchor la 00 76 3
    red) elict uc. may oncil me pir cied yere judge con is to chally.
   ENERS AND SCIUCY LEE 2. PRY ANJOHN JULION CHORD FORD ASID CONING ASIDES CON CERT SILVED MOUNTS FIND YOUNG JULION CERT SILVED MOUNTS FIND ASIDES
  שמוים און ליו פנסא דנותר לשת חוק שומר שימים דנוסת הבעון ובנכץ לכיוה מת
      הבתרון שי הנו ושנו- וות להכי הניתטור בנו ב הנוקטורי נגי להתחשב בדוראים תוכפים.
 - זום מוא תיפוט מופאר בנותרת יווי מתחשבי במספר בופאה שו באיל שנים, וך שיום
                                            had the mile allege some
        W term = 2. Max ( On and more) (pow)
                                    60767 -10 DILE 11(10)
                         האושר את השבו- -שבו מקכימוי לאיל ב דכותרת
                       عور الما إلى المولاد عوروس ، كي رورد مولاد عمر الماد المولاد
         Whern= 2
        יוי מוא בת פוש או מול זה לכנתרא אל מתחשבים ביצירו הונאה בחנמן ונשתנים
                 of the state often of the one to cult will and
         (Morn = J. Fij + B. MAR DR 2000)
                                                                PILLARY
                                                               1011
                                    og gister luchj.
                                                                      /P:0.3
                                                              150R
                                     一いれらっているいか
```

TNONT