

הוראות לנבחן בצידו השני של הדף

אין לכתוב מעבר לשוליים משני צידי הדף

שאלה 2 3

$$\frac{28}{32} - \frac{1}{30} = \frac{26}{30}$$

שאלה 5 9

שאלה 3 10 14

15

32

מס' כיתה 306 בנין

מס' נבחן 020258

20/02/2012 1345302 2 40/20  
הנדסת מערכות מידע  
אחזור מידע וספריות דיגיטליות  
03721440601101




המחלקה מ.מ.ר שנה 2009  
תאריך בחינה 26/2/12  
מקצוע בחינה מחשבי

**לשימוש המרצה הבודק**

יחידות ועשרות ומאות

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

ציון הבחינה 85  
שם המרצה ד"ר  
חתימה \_\_\_\_\_  
תאריך \_\_\_\_\_





אוניברסיטת בן-גוריון בנגב

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 לזרזי לשיחח רזמי הבחינה או לעזור אח מהומי ללא וטולח רשוח



איחזור מידע תשע"א – 372.1.4406

סמסטר חורף מועד א' 31.01.11

פרופ' ברכה שפירא, אורלי מורנו

משך המבחן : שעתיים וחצי

חומר עזר: מותר (לא מחשב נייד), מותר מחשבון

יש להחזיר את גיליון הבחינה. המבחן כולל 4 דפים

חלק א' - 30% - יש לענות על גבי הטופס

סמן תשובה אחת נכונה

1. 5% ההשפעה של חוק zipf באה לידי ביטוי באורך רשימות postings של terms על פי הפירוט הבא (רשימות posting הן הרשימות הכוללות אינפורמציה על מופעים של Terms במסמכים):

- א. אין קשר בין הדברים
- ב. רשימות של מעט מילים תהיינה מאוד ארוכות ושל הרבה מילים תהיינה מאוד קצרות
- ג. רשימות של הרבה מילים תהיינה מאוד ארוכות ושל מעט מילים תהיינה מאוד קצרות
- ד. החוק יבוא לידי ביטוי רק אם יופעל תהליך stemming על המסמכים
- ה. החוק יבוא לידי ביטוי רק אם יופעל תהליך של הסרת stopwords
- ו. ב+ד נכונים
- ז. ב+ה נכונים
- ח. ג+ה נכונים

2. 4% מדדי ההערכה הבאים מתאימים להערכת מנועי חיפוש אינטרנטיים:

- א. DCG , F-measure, precision@k , R-precision
- ב. DCG , precision@k , interpolated average precision , MAP
- ג. DCG , precision@k , MAP , Reciprocal Rank
- ד. MAP , precision@10 , Fallout , NDCG
- ה. כל המדדים שהוזכרו בהרצאות יכולים להתאים גם למנועי חיפוש אינטרנטיים
- ו. אף אחד מהמדדים שהוזכרו בהרצאות לא יכול להתאים בלי שינויים בהגדרות

3. 4% הפרמטר d בנוסחת Pagerank

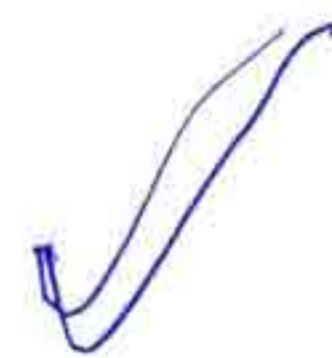
- א. מאפשר למדל באופן נכון את התנהגות ה Random surfer (הגולש האקראי) ולפתור את בעיית ה spider trap
- ב. מאפשר לחשב pagerank באופן מהיר
- ג. מאפשר לנרמל את תוצאת ה Pagerank
- ד. הוא הווקטור העצמי של מטריצת המעברים
- ה. מאפשר למשתמש לעבור לדף באופן אקראי
- ו. א+ה נכונים





4. 4% עדכון של האינדקס מיד עם הגעה של מסמך חדש מה crawler

- א. מאט את פעולת האינדוקס
- ב. משפר בהכרח את ה precision של שאליות רלוונטיות
- ג. משפר בהכרח את ה recall של שאליות רלוונטיות
- ד. מעלה את העדכניות של האינדקס אך מסבך את הניהול שלו
- ה. משפר את המהירות של ביצוע השאלות.



5. 5% נתון מאגר עם 4 המסמכים הבאים:

$D_1$ : alpha bravo charlie delta echo foxtrot golf

$D_2$ : golf golf golf delta alpha

$D_3$ : bravo charlie bravo echo foxtrot bravo

$D_4$ : foxtrot alpha alpha golf golf delta

לשאלת alpha bravo

- א. מנוע שמבוסס על vector space יחזיר את כל המסמכים ואת מסמך  $D_1$  במקום הראשון
- ב. מנוע שמתבסס על vector space יחזיר את כל המסמכים ואת מסמך  $D_3$  במקום הראשון
- ג. מנוע שמתבסס על מודל בוליאני טהור יחזיר את מסמך  $d_1$
- ד. מנוע שמתבסס על מנוע בוליאני טהור לא יחזיר את מסמך  $d_2$
- ה. ג+ד נכונים
- ו. ב+ג נכונים



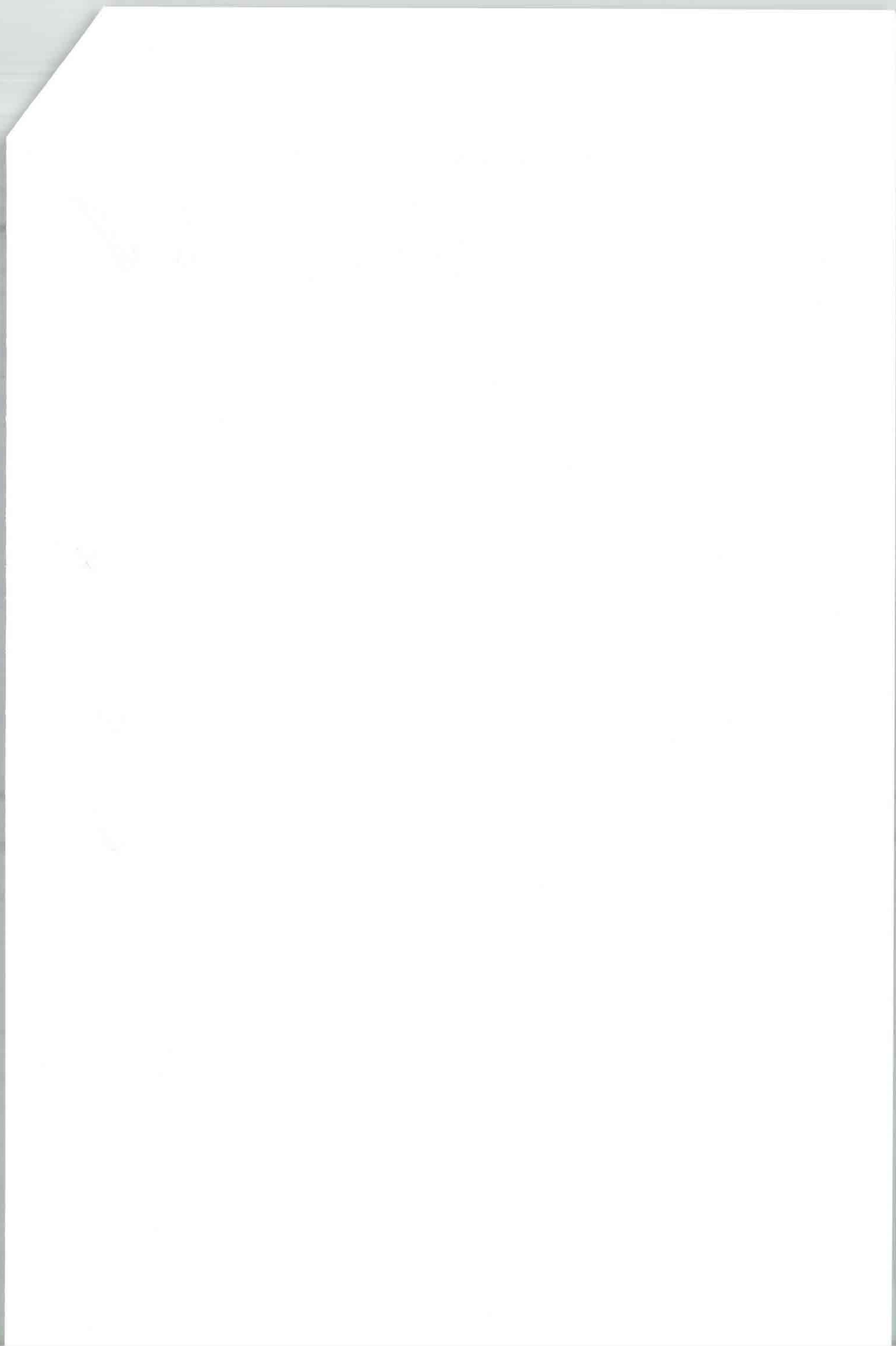
סמן נכון או לא נכון:

6. 4% כדי לאפשר מענה לשאליות של ביטויים האינדקס חייב להיות במבנה של biword index . נכון/לא נכון



7. 4% Relevance feedback יכול להתבצע באמצעות האלגוריתם של Rocchio אשר מעדכן את ווקטור השאלות שהמשתמש הקליד על פי תגובת המשתמש לתוצאות שהמנוע החזיר לשאלות. בהתאם לערכי הפרמטרים שקובעים את רמת ההתחשבות בשאלת המשתמש לעומת תגובתו לתוצאות השאלות. נכון/לא נכון







חלק ב' 70%

ענה על השאלות הבאות:

1. 30%

הנח שאילתא q1 שלה 5 מסמכים רלוונטים במאגר על פי הפירוט הבא,

מסמך d13 רלוונטי באופן מושלם (ציון 3 על סקאלה של 0-3 כאשר 0 הוא לא רלוונטי ו-3 רלוונטי באופן מושלם), מסמכים d10 ו d9 רלוונטים מאוד (ציון 2), מסמכים d1 ו d3 רלוונטים באופן סביר (ציון 1) וכל שאר המסמכים במאגר לא רלוונטים. (הערה: כל המסמכים שקיבלו ציון גבוה או שווה ל-1 נחשבים רלוונטים). מנוע E1 הפעיל את שאילתא q1 על המאגר הנ"ל והחזיר 20 מסמכים. ידועים רק 9 המסמכים הראשונים שהמנוע החזיר ואת הסדר שבו החזיר אותם (משמאל לימין):

d13, d2, d8, d15, d3, d10, d1, d9, d4.....

ענה על השאלות הבאות:

- א. 2% מהו ה precision ב 10 מסמכים של המנוע, e1 על פי השאילתא הנ"ל
- ב. 3% אם המנוע היה אידיאלי לשאילתא הנ"ל (כלומר היה מחזיר את כל המסמכים הרלוונטים ראשונים, מדורגים לפי רמת הרלוונטיות שלהם), מה היה אז ה precision ב 10 מסמכים?
- ג. 4% מהו ה r-precision, ומהו ה Reciprocal Rank של מנוע e1 על פי השאילתא הנ"ל
- ד. 4% מהו ה precision ב 30% recall?
- ה. 5% חשב DCG ב 5, השתמש ב discount של  $1/\log_2(\text{rank})$
- ו. 5% מנוע e1 הריץ שאילתא q2 שלה 15 מסמכים רלוונטים במאגר, המנוע החזיר 15 תוצאות על פי הסדר הבא (משמאל לימין, על כל מסמך שחזר מסומן אם הוא רלוונטי או לא, כאשר "ר" מסמן מסמך רלוונטי ו "ל" מסמן מסמך לא רלוונטי):  
ל,ר,ל,ל,ר, ל, ל,ל,ר, ל, ל,ל,ר, ל,ל,ר,ר,ר,

חשב את interpolated average precision על פי שתי השאילתאות.

- ז. 3% חשב MAP של המנוע על פי שאילתא q1.
- ח. 6% איך ישתנה ה MAP אם מתווסף מסמך רלוונטי נוסף למאגר והמנוע מחזיר אותו בין ה 20 המסמכים הראשונים. (הראה את ה MAP המקסימלי והמינימאלי האפשריים בעקבות השינוי).

2. 5% במערכת סינון שמקבלת באופן קבוע זרימה של מסמכים חדשים, לכל מסמך שמגיע למערכת מחושב הדמיון שלו עם פרופיל המשתמש. נניח שכל מסמך מיוצג כווקטור על פי גישת  $tf \cdot idf$  וכך גם פרופיל המשתמש. מהו idf במקרה זה, כלומר לאיזה מאגר אפשר לייחס את חישוב ה idf?

3. 15% לצורך תיקון שגיאות של שאילתאות, מחשבים מרחק בין מילת שאילתא שמשתמש הקליד ואיננה נמצאת באינדקס ומילים פוטנציאליות שאליהן התכוון המשתמש. הצע **נוסחא** לחישוב מרחק בין מילה שגויה לבין מילה פוטנציאלית מתוקנת שתתחשב במרחק בין אותיות במקלדת (בין אותיות שאולי הוחלפו בין המילים לתיקון) וכן בהפרש בין אורך המילה המקורית לאורך המילה המוצעת כתיקון. הראה באופן מפורט את כל חלקי הנוסחא הנדרשים, למשל הראה כיצד מחושב מרחק בין האותיות במקלדת. (כתשובה יש להראות נוסחה ולא הסברים מילוליים)







4. 10% קיימות שיטות שונות לקודד מילים על פי איך שהן נשמעות. השיטה הידועה ביותר היא שיטת ה Soundex – שהיא שיטה המבוססת על פונטיקה שהומצאה בשנת 1918. כל מילה מקודדת לאות שאחריה 3 ספרות. האות היא האות הראשונה של המילה, והספרות נקבעות על ידי כללים שונים שמקבצים כמה אותיות דומות לאותה ספרה. כך למשל, N i M מקודדות לסיפרה 5. V i , P , F , B מקודדות לסיפרה 1. אותיות ניקוד אינן מיוצגות, אלא אם כן האות הראשונה של המילה היא אות ניקוד. להלן כמה דוגמאות למילים מקודדות בשיטת soundex:

הקוד S-530 הוא הקוד של המילים: smythe-i smith  
הקוד a-450 הוא הקוד של המילים: Allan, Allen, Alan, Allyn ועוד.....  
Retrieve i retrieval מקודדות ל r-361 (משום שמתעלמים מאותיות עודפות מעבר לשלוש ספרות קוד).  
ציין שני יתרונות ושני חסרונות (שונים) לשימוש ב Soundex כשיטה לייצוג המילים באינדקס- במקום להשתמש ב Terms עצמם (או ב stem שלהם).

5. 10% מנועי חיפוש משתמשים בתוצאות של שאילתות זהות קודמות כדי לשפר תוצאות של שאילתא נוכחית. כלומר, המנועים שומרים במאגר מיוחד את השאילתות של המשתמשים, את התוצאות שהם החזירו וגם את התנהגות המשתמש עם התוצאות, כלומר אילו מהתשובות הוא אהב. כאשר משתמש מקליד שאילתא, המנוע מחפש במאגר המיוחד שאילתא זהה (אחת או יותר) ומשתמש בתוצאות של השאילתות הזוהות כדי לשפר את תוצאות השאילתא הנוכחית. הבעיה היא שכ 50% מהשאילתות שנשלחות למנועי חיפוש הן ייחודיות (כלומר, אין להן שאילתות זהות). הסבר כיצד אפשר להשתמש בתוצאות של שאילתות קודמות, גם כאשר השאילתא שהמשתמש הקליד אינה זהה לשאילתא שקיימת במאגר המיוחד. (אין צורך להציג אלגוריתם מדויק בפסדו-קוד אלא לספק הסבר ברור על השיטה שאתם מציעים)

בהצלחה - ברכה ואורלי







~~40/20~~  
40/20

חלק 2

שאלה 1

$$P@10 = \frac{5}{10} = 0.5$$

28  
38

(\*) חסר 5 בלויטם למ

(\*) סדר החסר החסרים את משיר R ה Precision 10-7  
חסרים, בלויטם ה Precision 10-7  
אם החסר היה אצל לחולן הוא היה חסר רק 5 חסרים

$$R\text{-Precision} = \frac{2}{5}$$

(\*) (2) חסרים בלויטם 5 חסרים

$$\text{Reciprocal Rank} = 1$$

(\*) (חסר כלן חסרים בלויטם)

(\*) חסרים חסרים חסרים 1 ה Recall חסרים  
חסרים חסרים חסרים 2 ה Recall חסרים

לחלק, precision חסרים חסרים חסרים

$$\text{Precision @ 0.3 Recall} = \frac{2}{5}$$

$$DCG = rel_1 + \sum_{i=1}^p \frac{rel_i}{log_2 i}$$

(\*)

Rank	1	2	3	4	5
rel <sub>i</sub>	3	0	0	0	1
Discount	3	$\frac{0}{log_2 2}$	$\frac{0}{log_2 3}$	$\frac{0}{log_2 4}$	$\frac{1}{log_2 5}$
DCG	3	3	3	3	3.43

$$DCG_5 = 3.43$$



$d_{13}, d_{14}, d_{15}, d_{16}, d_{17}, d_{18}, d_{19}, d_{20}$   
3 0 0 0 1 2 1 2 0







Mathematics



## שאלה 2

אם תישק ה-IDF אצל ליווי למאמר שגור לפסל  
השורה. למחר משה IDF של ~~המאמר~~

אם ככה מסתכן מנסה המילד מוזק ב המסתכן

הקוויט למעט הסכום הנה - מילד המסתכן האלה  
מאז זה יצוק להצדק (למחר משה) חישוק השקלים?

ב שני שטיל מסומן ב שני קוויט למעט

נצטן א ב המעלות ל השורה הסכום הנה

א/ק ב/א יתעלה!

לא כדור למחר זה נכח למעט קרי  
מסתכן וקוויט למעט

שאלה 3

-2

שקט שני מילד מוכח למחר Distance  
כך למחר קוויט

נכן משה מוזק ב מילד מוזק ב אור מוזק אור

מילד. למחר למחר A מוזק ל המוזק

Z, Q, W, S יהיה מוזק 1 - מוזק ל המוזק

D, C, F יהיה מוזק 2 וכך. (למחר מוזק)  
למחר מוזק למחר מוזק למחר מוזק

מילד למחר מילד

מוזק מוזק מוזק מוזק מוזק (השורה והאוסף למחר)

ומחר א מוזק מוזק למחר למחר למחר

(כמוק שגש הוא מוזק הוא אור מוזק הוא ס).

מוזק מוזק, מוזק ל המוזק שגש מוזק בין מוזק המוזק

מוזק מוזק מוזק המוזק המוזק מוזק מוזק.

למחר המוזק הוא: (מוזק א מוזק)

מוזק למחר A (ב ב א א) וא למחר למחר B

$$\text{Rank} = \left( \sum_{i=0}^{A.\text{length}} \text{Distance}[A[i], B[i]] \right) + |A.\text{length} - B.\text{length}|$$

כך למחר מוזק

אמחר מוזק א מוזק מוזק מוזק (מוזק מוזק)  
הנה למחר מוזק



Mathematics

Let  $x$  and  $y$  be real numbers such that  $x^2 + y^2 = 1$ .

Prove that  $x^2 \leq 1$  and  $y^2 \leq 1$ .  
Since  $x^2 + y^2 = 1$ , we have  $x^2 = 1 - y^2$ .  
Since  $y^2 \geq 0$ , it follows that  $x^2 \leq 1$ .  
Similarly,  $y^2 = 1 - x^2$ .  
Since  $x^2 \geq 0$ , it follows that  $y^2 \leq 1$ .



4 the

אברהם

(א) שמו של המלך יקין בן מנחם מלך יהודה  
(ב) אברהם ויחזקאל מלך ישראל  
ו' מלך ישראל ויחזקאל מלך ישראל.

(ב) יכלה הנצח מסתמך סמך שגור כג' אלה,  
 כח השגור כג' נכח אלה שגור א' א' א' א'  
 נח א' א' א' א' א' א' א' א' א' א'  
 כח השגור מסתמך אלה א' א' א'  
 א' א' א' א' א' א' א' א' א' א'

151907

(1) קהל מנהלים נפיק דיוור למספר, יוזמן מצבים שונים  
נפיק מנהל פונה אך אין עיכור לחומר נוסף  
פירוש (Smith - smite) ולכן כגד נצט  
לפניו מוזמן בליקור לקבל י' מצב שנה  
לפי מנהלים אין חלוקה.

(2) מיום פירוקה נאמרו 10218, ~~הוא~~  
 213' ב' פ'נ כ' זכור 'ה' 'ה'  
 נאמרו אחריו ארבעה אל תמהו .שפני)

↓  
אברהם אבינו  
זהו אברהם שבינת קדש

$$\frac{7.5}{10}$$



ALFRED H. ROYCE



## אלג' 5

נניח אנו השאלה באופן הבא הנתון, למנו  
 מכן ~~שאלה~~ יש מי איננו קוראים מאלו.  
 אנו הנתון ~~שאלה~~ שאלה יש מי איננו קוראים מאלו.  
 מיקום הנתון שאלה (שאלה הישגית, הנתון  
 (הנתון שאלה) אנו שאלה 5 הנתונים הנגזרים  
 מה נתון שאלה <sup>הנתון</sup> שאלה שאלה אנו הנתון הנתון  
 קוראים הנתון שאלה הנתון שאלה שאלה שאלה  
הנתון שאלה שאלה הנתון שאלה הנתון  
 קוראים שאלה שאלה שאלה שאלה שאלה

אנו שאלה שאלה שאלה שאלה שאלה  
 שאלה שאלה שאלה שאלה שאלה  
 שאלה שאלה שאלה שאלה שאלה

כדי למנו שאלה שאלה שאלה שאלה שאלה  
 שאלה שאלה שאלה שאלה שאלה  
 שאלה שאלה שאלה שאלה שאלה  
 שאלה שאלה שאלה שאלה שאלה

1. original results  $\Rightarrow$  הנתון שאלה
2. for doc 1 to 5 in original results
  - 2.1 for doc in best query for doc
    - 2.1.1 if doc exists in original query results
    - 2.1.2  $W[doc]$  in original  $\neq \infty$

-1

11

1. שאלה

2. for 1 to 5 most similar queries in database

...

\* שאלה שאלה שאלה שאלה שאלה  
 שאלה שאלה שאלה שאלה שאלה



Mathematics for Class 10

Chapter 1: Real Numbers

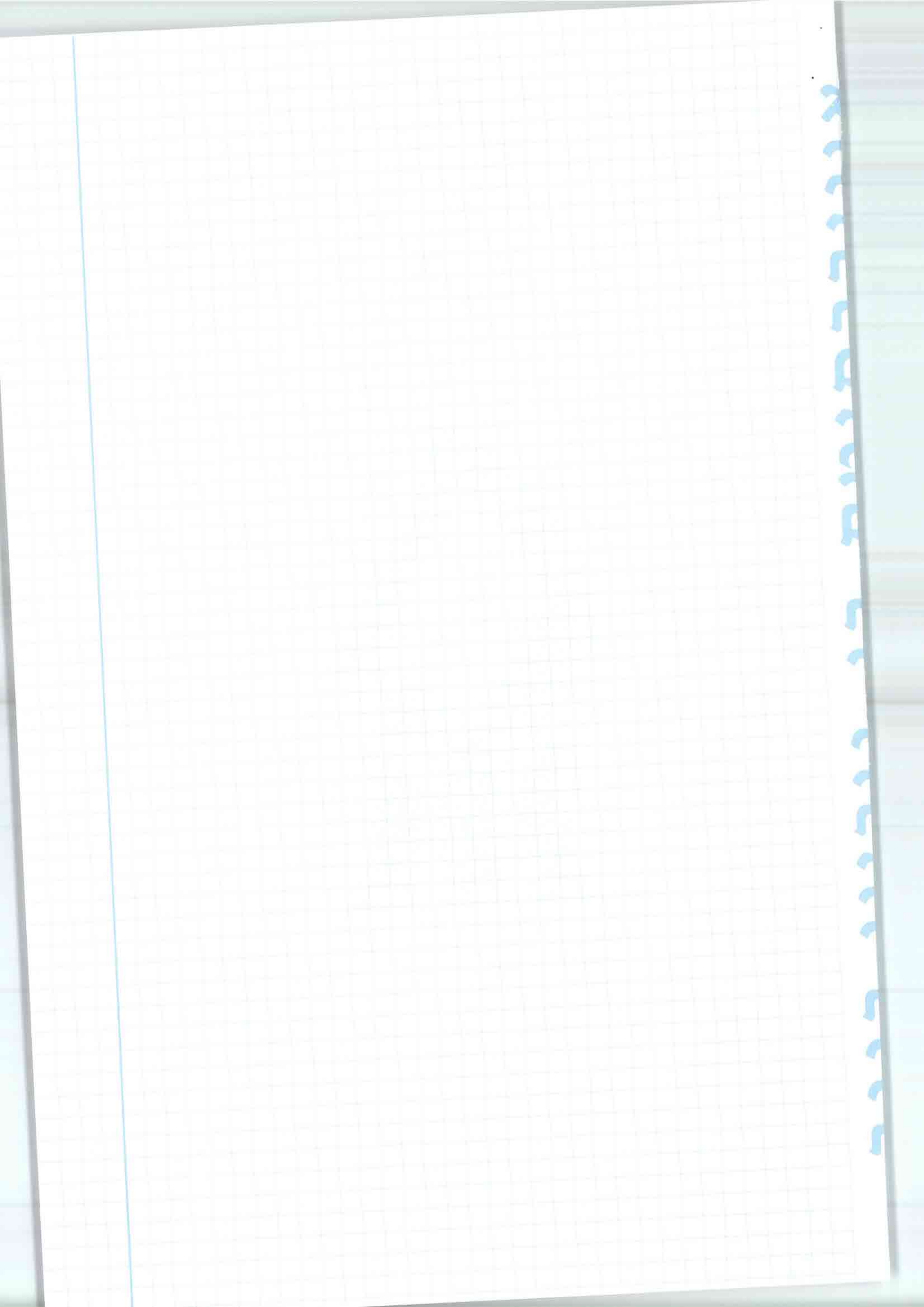


אנו מברכים את הנהנה



ALFRED H. ROSS, JR. / ALFRED H. ROSS, JR.







Mathematics



אברהם בן הרב



ALFRED H. ROYAL



אברהם בן יצחק



WILSON'S



