

הוראות לנבחן בצידו השני של הדף

אין לכתוב מעבר לשוליים משני צידי הדף

מס' כיתה \_\_\_\_\_ בנין \_\_\_\_\_

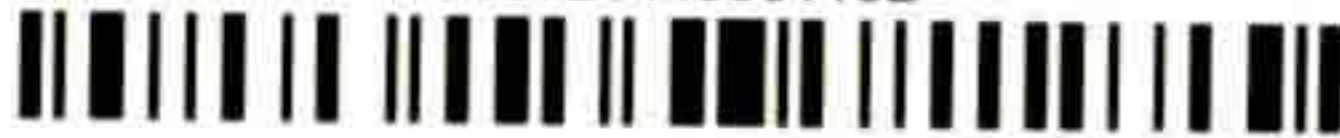
מס' נבחן **063602**

17/02/2011 1261520 2 32/12

הנדסת מערכות מידע

אחזור מידע וספריות דיגיטליות

03721440601102



### 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 \_\_\_\_\_

שם המרצה \_\_\_\_\_

חתימה \_\_\_\_\_

תאריך \_\_\_\_\_

המחלקה \_\_\_\_\_ שנה \_\_\_\_\_

תאריך בחינה \_\_\_\_\_

מקצוע בחינה \_\_\_\_\_



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 לנבחן לשוחח בזמן הבחינה, או לעזוב את מקומו ללא נטילת רשות.

הנבחרת לנבחן חייבת להגיע לפחות 15 דקות לפני שעת הנבחן. אין להחזיר השאלונים ואין



הוראות לנבחן בצידו השני של הדף

אין לכתוב מעבר לשוליים משני צידי הדף

חלק ז' - 9

32/12

שאלה 1 - 10  
שאלה 2 - 4  
שאלה 3 - 13

מס' כיתה 16 בנין 34

מס' נבחן 023085

מדבקה

1

מדבקה

המחלקה נציג מרכז שנה 17

תאריך בחינה 17/2/2011

מקצוע בחינה מתמטיקה מ-3

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 64

שם המרצה

חתימה קנדה שפירא

תאריך

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.**



איחזור מידע תשע"א – 372.1.4406  
סמסטר חורף מועד ב' 17.02.11  
ד"ר ברכה שפירא, אורלי מורנו

משך המבחן : שעתיים וחצי  
חומר עזר: מותר (לא מחשב נייד)  
יש להחזיר את השאלון

מס נבחן 093085

# חלק א – נא לענות במחברת הבחינה

1. 26% נתון log של שאילות שמשתמשים שלחו למנוע, ולכל שאילתא סט של 30 מסמכים שחזרו לשאילתא. כמו כן, נתון לכל שאילתא על אילו מסמכים מתוך אלו שחזרו המשתמש הסתכל. כדי לנתח את האיכות של המנוע צריך לזהות מתי המשתמש חיפש באותו נושא – כלומר מבין השאלות של המשתמש- אילו מהן היו בהקשר לאותו צורך מידע בהנחה שאין זיהוי של משתמש על השאילתא (לא ידוע אילו מהשאילות התבצעו על ידי אותו משתמש ברצף).

א. 18% הצע פיתרון (אלגוריתם בפסאודו קוד) שיקבץ את השאילות של משתמש ברצף ל session – כלומר לזהות באותו רצף את השאילות של משתמש אחד – המנסה למצוא מידע בנושא מסוים ומתי מתחיל session של משתמש אחר (או אותו משתמש בנושא שונה).

ב. 8% הסבר מהן המגבלות של הפתרון – כלומר באילו מקרים הוא לא יצליח לזהות את ה Session (אם אין כאלה מה טוב..... הסבר מדוע אין מגבלות)

2. 8% נניח crawler שעובד ב batch ושומר את ה Repository באופן מבוזר. הביזור מתבצע תוך כדי ה crawling כאשר ה crawler מוריד את הדף ואז שומר אותו בצומת לפני הטיפול שמתבצע על הדף ("טיפול" למשל הוצאת הלינקים וכו'). מנה שתי סיבות שבגללן עדיף ל crawler לצורך יעול עבודתו, לבזר את הנתונים בצמתים כך שבצומת אחת יישמרו כל הדפים של אתר מסוים – על פני ביזור של דפים שונים של אותו אתר בצמתים שונים. (הסיבות צריכות להיות קשורות ליעול עבודתו של ה crawler – לא לאינדוקס בהמשך).

3. 16% למאגר מסוים ואוסף שאילות התבצעה הערכה של מומחים על הרלוונטיות של כל מסמך לשאילתא (לצורכי הערכה של מנועים). ההערכה כללה 3 סוגי תוצאות (במקום ה 2 הסטנדרטיות): 0 – כאשר המסמך לא רלוונטי, 1 כאשר מסמך רלוונטי חלקית ו 2- כאשר המסמך רלוונטי לגמרי.

- א. 13% כתוב נוסחה מדויקת שמגדירה את מדד precision@k בהתאמה לתוצאות כאלו.  
- ב. 3% הדגם את המדד precision@5 שהגדרת על התוצאות הבאות (משמאל לימין):

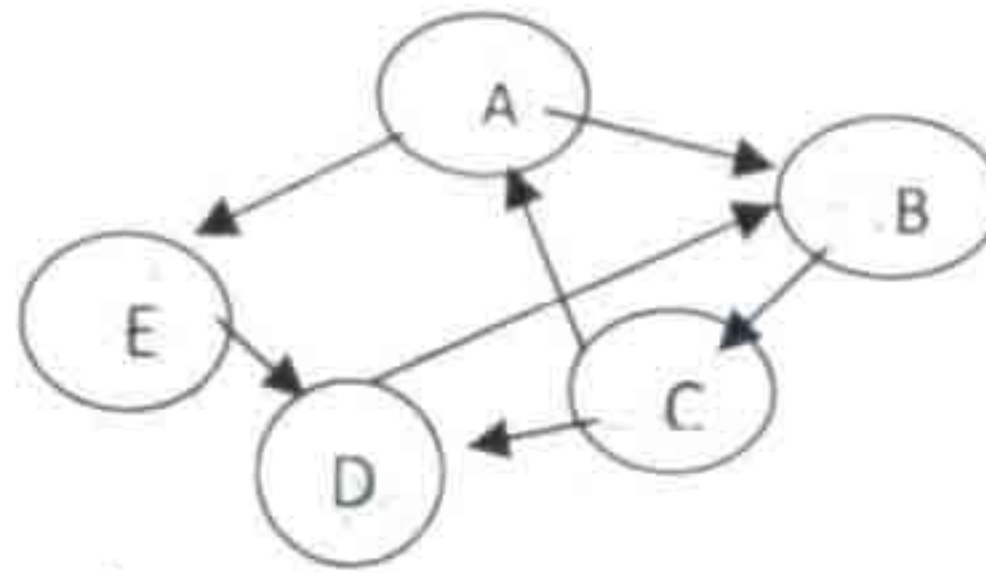
1,0,2,0,0,2,1,1  
1 2 3 4 5 6 7 8





חלק ב – יש לענות במחברת הבחינה במקומות המסומנים

4. 15% נתונה רשת :



השלם: (תשובה לא נכונה קונסת בחצי נקודה – כלומר על תשובה לא נכונה יופחתו 4.5 נקודות).

5% הצומת בעלת ערך ה hub הגבוה ביותר על פי אלגוריתם HITS הוא: C

5% בין הצמתים E, B, D הצומת בעלת ערך authority נמוך ביותר על פי אלגוריתם HITS היא E

5% 2 הצמתים בעלי ה pagerank הגבוה ביותר הם C, B

סמן נכון או לא נכון

5. 17%

- א. 3% ערך  $tf*idf$  של term במסמך יכול להיות גבוה יותר מ- 1. נכון/לא נכון ✓
- ב. 3% שימוש ב Stemmer יכול להשפיע על שביעות רצון המשתמש מהמנוע. נכון/לא נכון ✓
- ג. 3% צומת בגרף מסוים שהיא בעלת ה pagerank הגבוה ביותר בגרף, בהכרח תהיה גם בעלת ה Authority הגבוה ביותר. נכון/לא נכון. ✓
- ד. 4% הנח שתי מילים "teach" ו "taught" ששקללו אותן כשתי מילים נפרדות במסמכים בהן הן הופיעו, על פי  $tf*idf$  סטנדרטי. לאחר מכן הסתבר ששתי המילים הן בעצם stem של אותה מילה ושאפשר לייצג אותן במשותף כמילה אחת (בהנחה שהמנוע תומך ב stem). כדי לתקן את הטעות, צריך לסכום את ה  $tf*idf$  שלהן לייצוג המשותף. נכון/לא נכון ✓

ה. 4% נתון פרופיל משתמש לאחר עדכון על פי תגובת משתמש על פי אלגוריתם rocchio לאחר ניתוח תגובתו למסמכים d1 d2 d3. בהנחה שאין הפחתה שלילית (כלומר  $\gamma=0$ ). אפשר לראות שהמשתמש העדיף פוליטיקה ומדע על פני מוזיקה. נכון/לא נכון \*

ספורט	מוזיקה	פוליטיקה	מדע	רמטכ"ל	לימודים
0.3	0.2	0.8	0.8	0	0
0.2	0	0.8	0.9	0	0
0	0.9	0.2	0	0.2	0.1
0.4	0	0.8	1	0	0
מסמך d1					
מסמך d2					
מסמך d3					

6. 6% נתון מאגר ובו 4 מסמכים :

D1- Tibet Tibet Malaga <sup>max 2</sup>

D2- Malaga Rimini Salvador Tibet Tibet Tibet Tibet <sup>max 4</sup>

D3- mexico Sun 1

D4- Mexico Malaga Tibet Sun 1

נתונים שלושה מנועים – (E1) עובד לפי המודל הבוליאני הטהור, (E2) לפי המודל הווקטורי, (E3) לפי מודל בוליאני מורחב שמדרג את המסמכים לאחר הפעלת האופרטורים הבוליאניים בשאלתא.

השאלתא Malaga and Tibet נשלחה ל E1 ול E3

השאלתא Malaga Tibet נשלחה ל E2

- א. 3% E1 ו E2 יחזירו תשובה זהה לשאלתות שנשלחו אליהם. נכון/לא נכון ✓
- ב. 3% E1 ו E3 יחזירו את אותם מסמכים לשאלתא שנשלחה אליהם. נכון/לא נכון ✓







12% סמן תשובה אחת נכונה לשאלות הבאות

7. 4% מנועים מסוימים משתמשים בתגובות משתמשים לתשובות על שאלתא q1 ומריצים בעקבותיה שאלתא נוספת q2. (relevance feedback). שיטה זו:

- א. משפרת תמיד את ה Recall של שאלתא q1
- ב. עשויה לשפר את ה precision של שאלתא q1
- ג. אינה משפיעה על ביצועי המנוע אלא רק משפרת את שביעות רצון המשתמשים
- ד. משפרת precision ו recall של שאלתא q2

השיטה משפרת את הביצועים של המנוע.

-4

8. 4% לשאלתא Q 4 מסמכים רלוונטים במאגר. להלן תוצאות שחזרו משני מנועים: E1 ו-E2.

התוצאות משמאל לימין כשאר N מסמל מסמך לא רלוונטי, R מסמן מסמך רלוונטי:

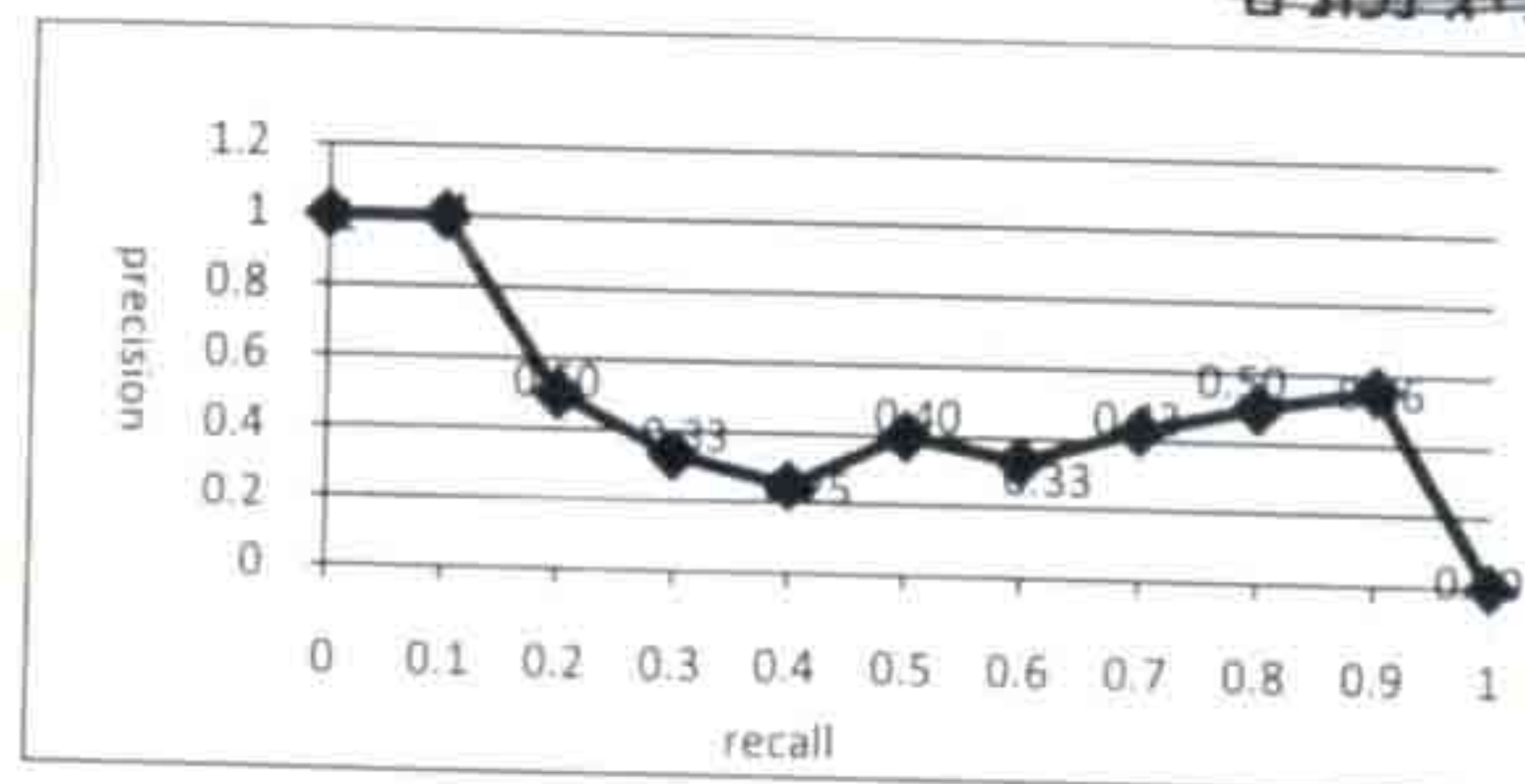
- E1: R N R N N N N N R R
- E2: N R N N R R R N N N

- א. ל E1 MAP גבוה יותר מאשר ל E2
- ב. R-precision של שני המנועים שווה
- ג. R-precision של E1 גבוה מ R-precision של E2
- ד. MAP של E2 גבוה מ MAP של E1
- ה. א+ג נכונים
- ו. א+ב נכונים
- ז. ג+ד נכונים

9. 4% נתון גרף precision-recall. הגרף משקף תוצאות של שאלתא אחת.

- א. הגרף התקבל בהכרח מנתונים אמיתיים ללא אינטרפולציה
- ב. הגרף התקבל בהכרח לאחר אינטרפולציה של ערכים
- ג. כל המסמכים הרלוונטים לשאלתא חזרו.
- ד. 90% לפחות מהמסמכים הרלוונטים לשאלתא חזרו.

- ה. ב+ד נכונים
- ו. ב+ג נכונים



בהצלחה

ברכה ואורלי





# אנליזה מתמטית - חלק 1 - פונקציות

ХИТРА ОВРАТ-ХИТРА





ХИТАЙСКО-КАРАКАУМСКАЯ



## ② מציאת הפתרון :

- במידה ולא תמצאו הקטנה למחצית הנורמלית של  $\cos$  במאגר המידות של נאליס, תוכלו להשתמש ב-  $\cos$  כמחצית של הקטנה הנורמלית של  $\cos$  וזה יהיה הפתרון הנדרש.

- ייתכן שיש מספרים שונים של  $\cos$  וזה יהיה הפתרון הנדרש. אם יש מספרים שונים של  $\cos$  וזה יהיה הפתרון הנדרש.

- אם  $\cos$  הוא הקטנה הנורמלית של  $\cos$  וזה יהיה הפתרון הנדרש. אם יש מספרים שונים של  $\cos$  וזה יהיה הפתרון הנדרש.

לכן לפי א'

2610

לשקב, 10

זמן זמן	Recall		Prec	
	ה	ע	ה	ע
1	x	$\frac{0}{10} = 0$	$\frac{0}{1} = 0$	0
2	1	$\frac{1}{10}$	$\frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$
3	2	$\frac{2}{10}$	$\frac{2}{3}$	$\frac{2}{3} \cdot \frac{1}{2} = \frac{2}{6} = \frac{1}{3}$
4	0	$\frac{0}{10}$	$\frac{0}{4}$	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$
5	2	$\frac{2}{10}$	$\frac{2}{5}$	$\frac{2}{5} \cdot \frac{1}{2} = \frac{2}{10}$
6	1	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{2}{3} \cdot \frac{1}{2} = \frac{2}{6} = \frac{1}{3}$
:				

1177

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\frac{2}{3}$$

$$\frac{2}{4} \cdot \frac{1}{3} = \frac{2}{12} = \frac{1}{6}$$

$$\frac{3}{5}$$

$$\frac{2}{3} \cdot \frac{1}{2} = \frac{2}{6} = \frac{1}{3}$$

20

2 - האותיות הקטנות - נכנסות

1 - (האותיות) - נכנסות -  $\frac{1}{2}$

0 - נכנסות -  $\frac{1}{3}$



3 - Rec

3 - Rec

precision @ k = the prec. at Doc Number k

3 - Rec (K)

$$\text{precision @ } k = \begin{cases} \text{old prec} * \frac{1}{3} \\ \text{old prec} * \frac{1}{2} \\ \text{old prec} \end{cases}$$

at Doc # 0 ~~0~~

at Doc # 1

at Doc # 2

(-10)

1/10/10

Doc #	Rec	old prec	new prec @
1		$\frac{1}{1} = 1$	$1 \cdot \frac{1}{2} = \frac{1}{2}$
2		$\frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
3		$\frac{2}{3}$	$\frac{2}{3} = \frac{2}{3}$
4		$\frac{2}{4}$	$\frac{2}{4} \cdot \frac{1}{3} = \frac{2}{12} = \frac{1}{6}$
5		$\frac{2}{5}$	$\frac{2}{5} \cdot \frac{1}{3} = \frac{2}{15} = \frac{2}{15}$
6		$\frac{1}{6}$	$\frac{1}{2} = \frac{1}{2}$
7		$\frac{1}{7}$	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$
8		$\frac{5}{8}$	$\frac{5}{8} \cdot \frac{1}{2} = \frac{5}{16}$

(2)

(-3)

precision @  
1/10/10



2610

(K)

$$H(D) = A(B)$$

$$H(E) = A(D)$$

11 D

$$B = C = A + D$$

$$\begin{array}{c} C > A > B \\ \hline C > D \\ \hline \end{array}$$

$B > E$

$$B > E$$

$$0 > A$$

$\Gamma, B, D$

IN A 75%?

DD (E) sk 'og N H(C)

$$O \sin E$$

~~P. 12~~

Q A 3

$C = B$

$$\textcircled{B} > A$$

$$A > E$$

$$B > D > \cancel{E}$$

$$C = B$$

$\odot > \Xi$

$$D = \frac{3}{4}(C)$$





2- t f i d f

2016

6

idf t f / 110

E1

	Tibet	Malaga	Rimini	Salvador	Mexico	Sun
D1	1	1				
2	$\frac{2}{2} \cdot \log \frac{4}{3} = 0.41$	$\frac{1}{2} \log \frac{4}{3} = 0.2$				
3						
D2	1	1	1	1		
2	$\frac{4}{4} \log \frac{4}{3} = 0.41$	$\frac{1}{4} \log \frac{4}{3} = 0.1$	$\frac{1}{4} \log \frac{4}{1} = 0.2$	$\frac{1}{4} \log \frac{4}{1} = 0.2$		
3						
D3	1				1	1
2					$\frac{1}{1} \log \frac{4}{2} = 1$	$\frac{1}{1} \log \frac{4}{2} = 1$
3						
D4	1	1			1	1
2	$1 \log \frac{4}{3} = 0.41$	0.41			$1 \log 2 = 1$	1
3						

250 - E2 - Malaga Tibet

$$\text{Sim}(D4) \quad 0.41 + 0.41 = 0.82$$

D4  
D1  
D2

D1  
D2  
D4

E1 - Malaga Tibet

E3 -

Yours ink - 1101 2



20 סיכורין אין :

1) סיכורין הוואנדען נאכער נאכער שטענדיג - crawler

2) באשענען דעמאלט אין דעם (באשענען) און אונדזערע,   
 3) דעם דעם אונדזערע אונדזערע "סיכורין" אונדזערע   
 4) דעם דעם אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע

2) סיכורין נאכער דעם אונדזערע אונדזערע אונדזערע

סיכורין דעם אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע

אונדזערע אונדזערע אונדזערע אונדזערע



2010

8

4 אינוניס"ם דאטא

גודל	Rec E1	Pre
1	$\frac{1}{4} *$	$\frac{1}{1} = 1$
2	$\frac{1}{4}$	$\frac{1}{2}$
3	$\frac{2}{4} - \frac{1}{2} *$	$\frac{2}{3}$
4	$\frac{2}{4} - \frac{1}{2}$	$\frac{1}{2}$
5	-	$\frac{2}{5}$
6	-	-
7	-	-
8	-	$\frac{2}{8}$
9	$\frac{2}{4} *$	$\frac{3}{9}$
10	$\frac{4}{4} = 1 *$	$\frac{4}{10}$

גודל	Rec E2	Pre
1	$\frac{0}{4} = 0$	0
2	$\frac{1}{4} *$	$\frac{1}{2}$
3	$\frac{1}{4}$	$\frac{1}{3}$
4	$\frac{1}{4}$	$\frac{1}{4}$
5	$\frac{2}{4} *$	$\frac{2}{5}$
6	$\frac{3}{4} *$	$\frac{3}{6} = \frac{1}{2}$
7	$\frac{4}{4} = 1 *$	$\frac{4}{7}$
8	$\frac{4}{4} = 1$	$\frac{4}{8}$
9	1	$\frac{4}{9}$
10	1	$\frac{4}{10}$

$$Map = \frac{1 + \frac{2}{3} + \frac{3}{9} + \frac{4}{10}}{4} = \frac{3}{5} = 0.6$$

$$\frac{\frac{1}{2} + \frac{2}{5} + \frac{1}{2} + \frac{4}{7}}{4} = 0.49$$

E1 > E2

2



# צונו במסר מטת בזז- גורר ליוז בזגב

ג.ר.ר - 10316 סידן בלחכת קסין

3

7

8

8

9

9

—

אברהם שמואל זאב-לוי



# צורכי משימה בזמן-גורמיוז בנוגב

# שאלות ותשובות - סדרה ב' - אורח



# אנליזה מסתמית - כד - גורמים כוללים

ХИТТИ ОҚИР-ХИЛИ ТҲТ



# אורליסלמזם כז-גורלויז כולגה

ХИТРА ОВОЌ И-ХИЛИТ





