

חלק א' - $\frac{4}{30}$ - 26
 חלק ב' -
 שאלה 1 - $\frac{19}{20}$ - 1
 שאלה 2 - $\frac{20}{20}$

הוראות לנבחן בצידו השני של הדף

אין לכתוב מעבר לשוליים משני צידי הדף

מס' כיתה 242 בנין 90

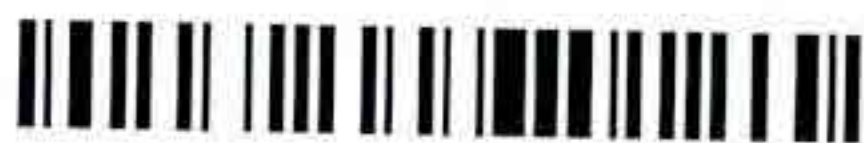
מס' נבחן 0088100

25/03/2012 1352181 1 30/13

הנדסת מערכות מידע

אחזור מידע וספריות דיגיטליות

03721440601402



לשימוש המרצה הבודק

יחידות ועשרות ומאות

<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
-1	1	4	3
-8	<input type="checkbox"/>	7	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

64

ציון הבחינה

למנהל

שם המרצה

חתימה

תאריך

המחלקה ^{הגם קוד} שנה

תאריך בחינה 25.3.12

מקצוע בחינה יחיד



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.

30/13

בס"ד

איחזור מידע תשע"א – 372.1.4406

סמסטר חורף מועד ב' 23.03.2012

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. יש לענות על כל השאלות

חלק א 30% - יש לענות על הגליון

בחר תשובה אחת נכונה

u
30

α, β, γ

1. 4% הפרמטרים בנוסחת rocchio ל relevance feedback :

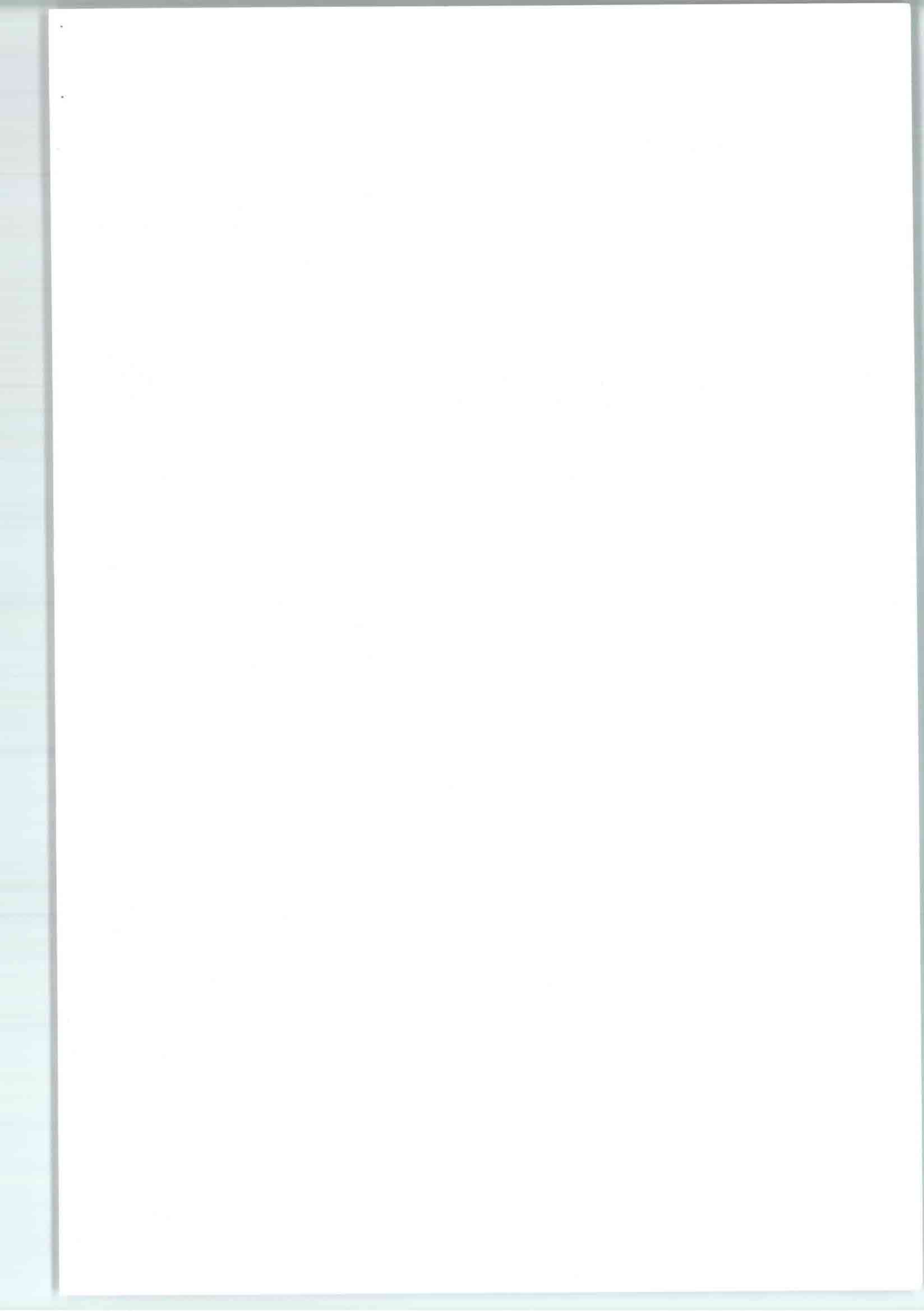
- ☒ א. קובעים האם המשתמש סימן את המסמכים כרלוונטים או לא.
☒ ב. נקבעים דינמית לפי תגובות המשתמש למסמכים
☒ ג. צריכים להסתכם ל - 1
☒ ד. קובעים את רמת ההתחשבות בשאילתא הנוכחית, במסמכים הרלוונטים ובמסמכים הלא רלוונטים בעדכון השאילתא
☒ ה. ג+ד נכונים
☒ ו. ב+ג נכונים

2. 6% נתונה טבלת הדירוגים הבאה של משתמשים על מוצרים בדירוג 1-5:

	User 1	User 2	User 3	User 4	User 5
Item1	1			1	
Item2		2	3		3
Item3	1	1	5	1	
Item4				4	
Item5	5		1		5

כדי לנבא דירוג של item5 עבור User4 על פי שיטת הדמיון בין המשתמשים:

- ☒ א. יש לחשב דמיון בין user4 לכל אחד מהמשתמשים האחרים ולחשב את הדירוג על פי המשתמשים הדומים ביותר
☒ ב. יש לחשב דמיון בין user4 ל user3 ו user1 ולהשתמש בדירוגים של שניהם ל item5 כדי לחשב את הדירוג של item5 ל user4
☒ ג. יש לחשב דמיון בין user4 ל user3 ו user1 ולהשתמש רק בדירוג של user1 ל item5 כדי לחשב את הדירוג של item5 ל user4
☒ ד. יש לחשב דמיון בין user4 ל user3 ו user1 ולהשתמש רק בדירוג של user3 ל item5 כדי לחשב את הדירוג של item5 ל user4
☒ ה. על סמך הנתונים הקיימים אי אפשר לנבא דירוג של item5 עבור User4
☒ ו. יש לחשב דמיון בין user4 ל user 5 ובין user 1 ו user 3



3. 5% המשקל של term בווקטור של מסמך במודל הווקטורי
- א. מבטא את מידת הייצוגיות של ה term למסמך
 - ב. צריך להיות מחושב באותו אופן שבו מחשבים את משקל ה Terms בשאלתא
 - ג. מחושב תמיד על פי $tf \cdot idf$
 - ד. חייב להיות בין 0-1
 - ה. ב+ג נכונים
 - ו. ג+ד נכונים
 - ז. א+ד נכונים

4. Benchmark 4%
- א. מגדיר מדדי ביצוע
 - ב. משמש להשוואה של ביצועי אלגוריתמים שונים מול מדדים חדשים שהוגדרו
 - ג. משמש להשוואה של ביצועי אלגוריתמים שונים מול תוצאות נכונות ידועות
 - ד. א+ג נכונים
 - ה. מחזק את הקשר בין האקדמיה לתעשייה
 - ו. א+ג+ה נכונים

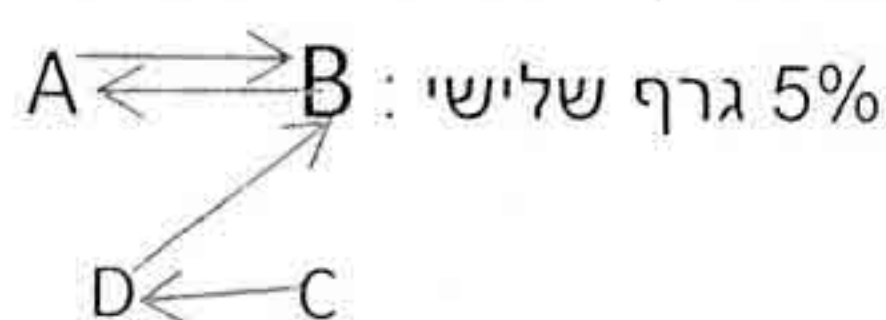
5. 3% - ההחלטה אם להתחשב ב Robots.txt תלויה ברמת האתיקה של אלגוריתם ה Crawling
- נכון/לא נכון

6. 4% DCG הוא מדד לבדיקת יכולת של מנוע לדרג גבוה דפים רלוונטים, תוך התחשבות ברמת הרלוונטיות של הדף
- נכון/לא נכון

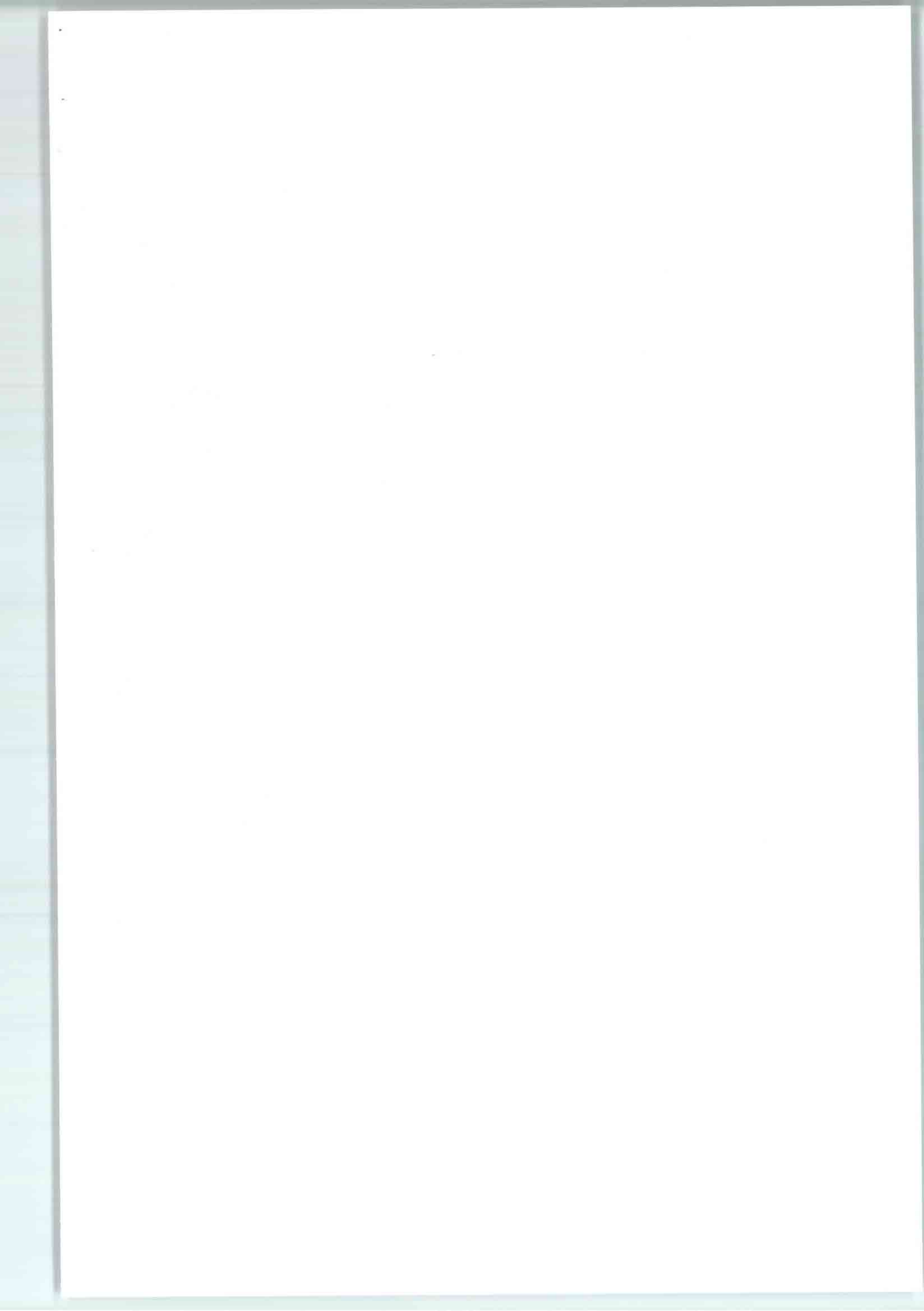
7. 4% stemming הוא תהליך תמיד בכללי שפה ולכן יש צורך להגדיר לכל שפה stemmer נפרד.
- נכון/לא נכון

חלק ב 70%

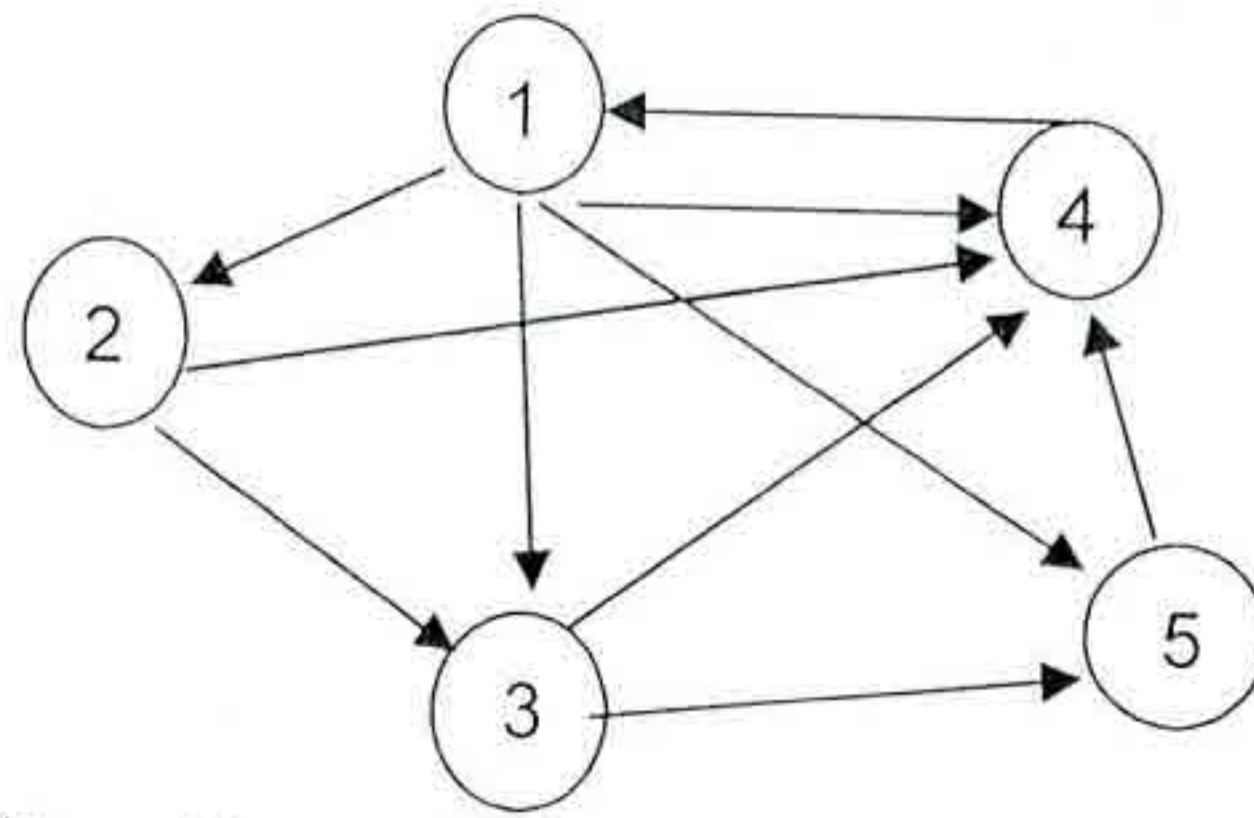
1. link analysis 20%
- א. 11% נתונים 3 גרפים שונים המתארים קישוריות של צמתים ברשת:
 - 4% גרף אחד: $A \rightarrow B \rightarrow C \rightarrow D$
 - 5% גרף שני: כמו הגרף הראשון, רק שבנוסף D מצביע על A (כלומר נוצר מעגל)
 - 5% גרף שלישי: $A \leftrightarrow B$



עבור כל אחד מהגרפים הסבר מה יהיו ערכי pagerank של כל הצמתים (אין לחשב ערכים מדויקים, אלא להעריך ולהסביר את היחסים בין ערכי כל הצמתים)

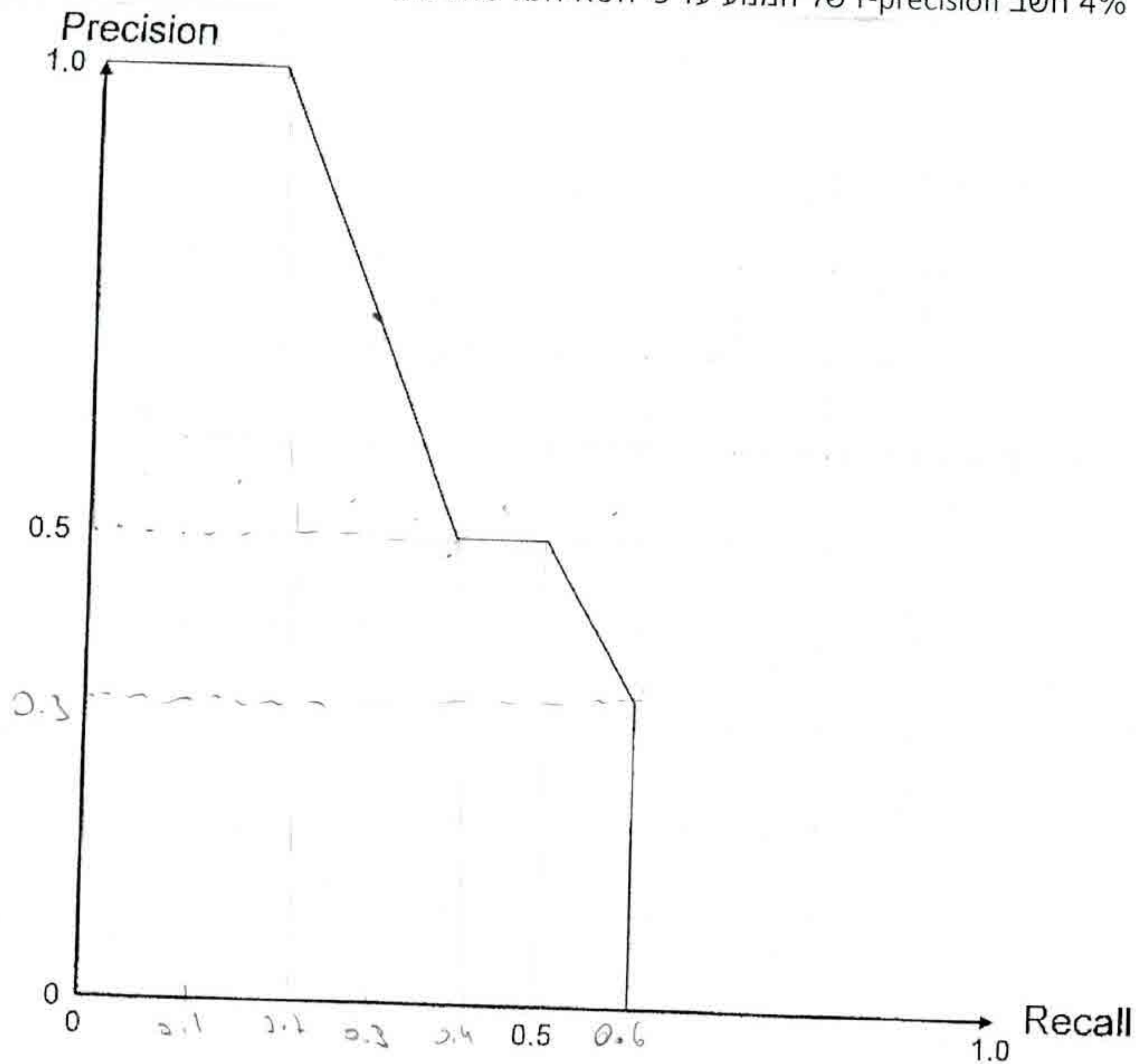


ב. 6% נתון הגרף הבא:



הראה את מטריצת המעברים (transition matrix) שעל פיה יחושב Pagerank לדפים (ללא התחשבות בפרמטר d).

2. 20% נתון גרף Precision-recall על תוצאות של הרצת מנוע מסויים על שאילתא מסויימת. המנוע החזיר 20 תוצאות. ידוע שבמאגר 10 מסמכים רלוונטים לשאילתא.
- א. 5. מהו ה precision לאחר שהמנוע החזיר 4 מסמכים רלוונטים.
- ב. 8% הראה את הרשימה שהמנוע החזיר (הצג רשימה של 20 מסמכים מדורגים וסמן מי מהם רלוונטי ומי לא)
- ג. 4% חשב את ה map של המנוע על פי השאילתא שהורצה.
- ד. 4% חשב r-precision של המנוע על פי השאילתא שהורצה.



3. 15%

א. 10% הועלה רעיון להשתמש ב jaccard similarity לצורך דירוג מסמכים לשאילתא (כלומר, לדרג את המסמך על פי מספר המילים המשותפות בין שאילתא ומסמך מחולק בסכום של מספר המילים של השאילתא והמסמך). אפשר להניח שכל המילים בשאילתא ייחודיות (אין חזרה על מילים) ושבמסמך סופרים כל מילה רק פעם אחת. באופן פורמאלי הדרוג למסמך D נקבע על פי:

$$jaccard(Q,D) = \frac{|Q \cap D|}{|Q \cup D|}$$

כאשר Q הוא מספר המילים הייחודיות בשאילתא, ו D הוא מספר המילים הייחודיות במסמך. תן 2 סיבות לכך שהרעיון אינו מוצלח, כלומר שתי סיבות לכך שהשיטה אינה מדרגת טוב מסמכים.

אם מוסיף
דבר "א"ע
לגוף המילה
השאלה

ב. 5% הדגם את חסרונה של השיטה מסעיף א. כלומר, תן דוגמא ספציפית לשאילתא Q כלשהי ושני מסמכים $D1$ ו $D2$. כאשר ברור ש $D1$ רלוונטי לשאילתא ו $D2$ לא רלוונטי. בחישוב על פי jaccard יוצא ש $D2$ יותר רלוונטי מ $D1$.

4. 15% בניסיון לשפר תוצאות של מנוע שעובד לפי השיטה הווקטורית, הוצע להוסיף לשאילתא מילים נרדפות למילות השאילתא מתוך מילון. לכל מילה בשאילתא הוצע להוסיף עד 10 מילים נרדפות. למשל: אם השאילתא היא: אחזור מידע – המילים הנרדפות לאחזור במילון הן: דלייה, אשוב, אשנה. מילים נרדפות למידע הן: ידיעה, אינפורמציה. השאילתא שתשלח למנוע תהיה: אחזור, דלייה, אשוב, אשנה, מידע, ידיעה, אינפורמציה

א. 8% תן סיבה אחת לסיכוי שהשיטה תשפר תוצאות וסיבה אחת לסיכוי שהשיטה תפגע באיכות התוצאות

ב. 7% הצע רעיון לשיפור השיטה שתגדיל את סיכוייה לשפר את התוצאות, הסבר את ההיגיון ברעיון שאתה מציע.

בהצלחה

ברכה ואורלי

30/13

אנו מודים לך על שיתוף

$\lambda C/C$

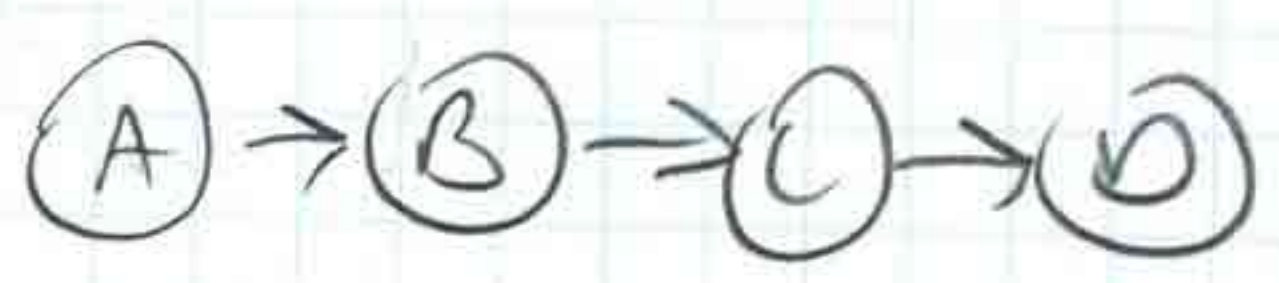


odun

P₁

	1	2	3	4	5
1	0	0	0	1	0
2	$\frac{1}{4}$	0	0	0	0
3	$\frac{1}{4}$	$\frac{1}{2}$	0	0	0
4	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	0	1
5	$\frac{1}{4}$	0	$\frac{1}{2}$	0	0

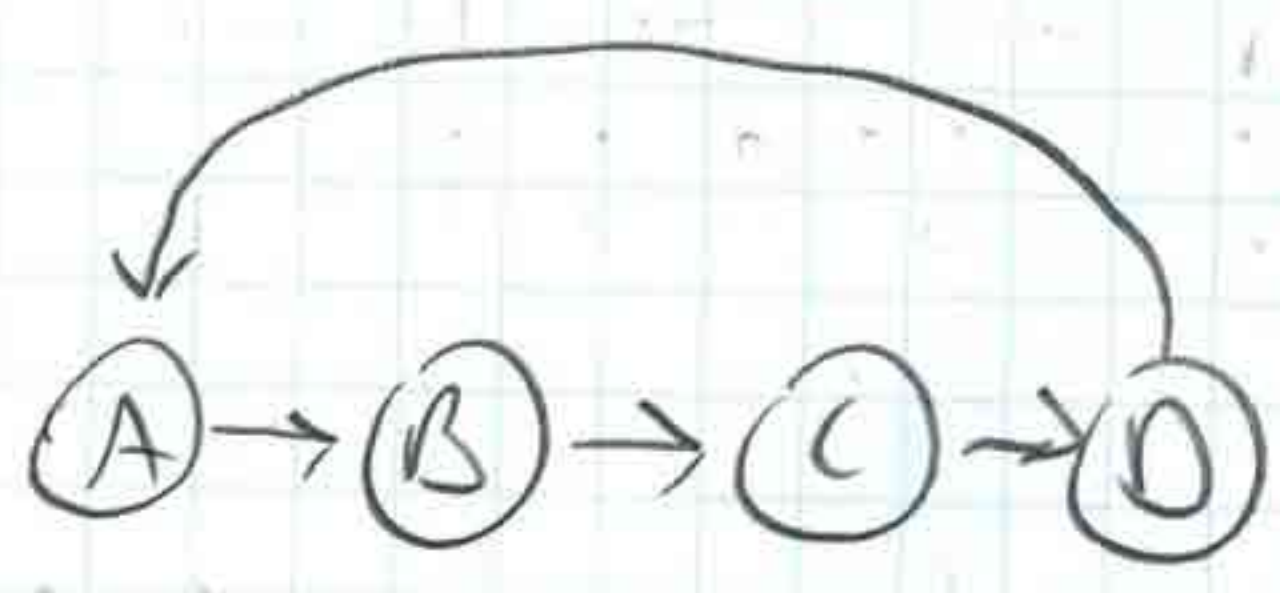
①
②



התחלה

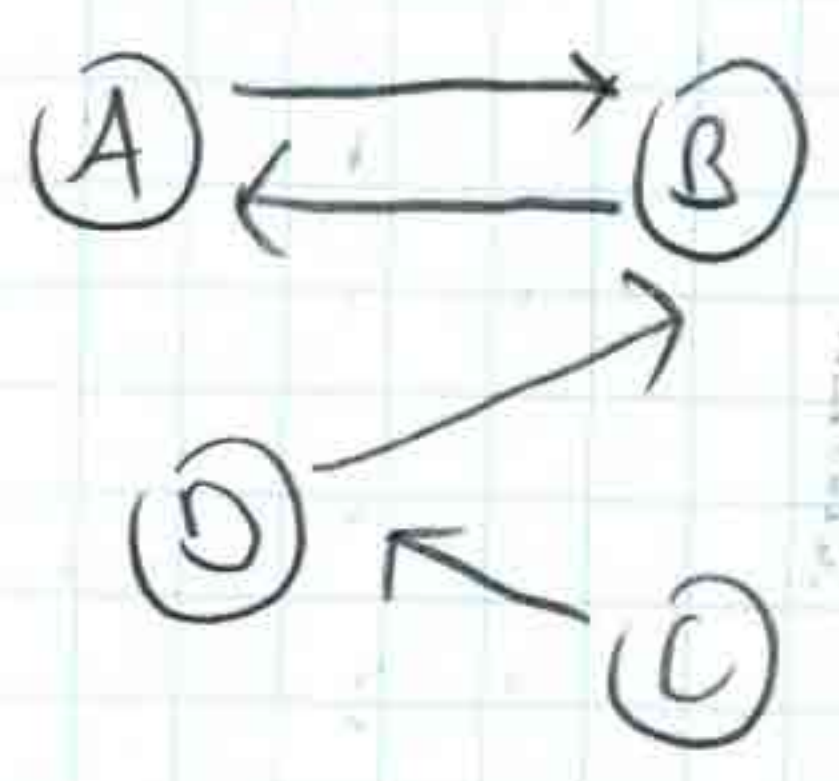
① > Dead end / מוצא כיוון e מוצא PR

כאשר ק"מ של מוצא (אולי) מוצא של מוצא
ואולי מוצא של מוצא (אולי) מוצא של מוצא.



התחלה

ה PR יוצא מוצא גרר מוצא. כי מוצא מוצא
גרר מ-1 ומוצא מוצא. כי PR מוצא מוצא.



התחלה

A ו B יוצא מוצא Spider trap כי מוצא מוצא
מוצא מוצא. כי מוצא מוצא מוצא מוצא
מוצא מוצא מוצא מוצא מוצא מוצא

מוצא מוצא. ומוצא מוצא.

	1	2	3	4	5
1	0	0	0	1	0
2	$\frac{1}{4}$	0	0	0	0
3	$\frac{1}{4}$	$\frac{1}{2}$	0	0	0
4	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	0	1
5	$\frac{1}{4}$	0	$\frac{1}{2}$	0	0

1/9
2/0

③

AGIC

$$P = \frac{1}{20}$$

$$R = \frac{1}{10}$$

$$0.3 \cdot 10 = \boxed{3}$$



	R	P
✓ 1	$\frac{1}{10}$	1
✓ 2	$\frac{2}{10}$	1 $\frac{1}{2}$
✓ 3	$\frac{3}{10}$	$\frac{3}{5}$
✓ 4		
✓ 5		
✓ 6		
✓ 7		
✓ 8	$\frac{5}{10}$	$\frac{5}{8}$
✓ 9		
✓ 10	$\frac{5}{10}$	$\frac{5}{10}$
11		
12		
13		
14		
15		
16		
✓ 17	$\frac{6}{10}$	$\frac{6}{20}$

$$\frac{0.3}{x} = p$$

$$x = 0.5$$

(1) Precision ≥ 0.5 (1)
 (2) Precision ≥ 0.4 (2)
 (3) Precision ≥ 0.3 (3)

R - relevant
 N - not relevant

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
 R R N R N N N R N R N N N N N N N N R

	R	P
✓ 1	$\frac{1}{10}$	1
✓ 2	$\frac{2}{10}$	1
3		
✓ 4	$\frac{3}{10}$	$\frac{3}{4}$
5		
6		
7		
✓ 8	$\frac{4}{10}$	$\frac{4}{8}$
9		
✓ 10	$\frac{5}{10}$	$\frac{5}{10}$
11		
12		
13		
14		
15		
16		
17		
18		
19		
✓ 20	$\frac{6}{10}$	$\frac{6}{20}$

✓ \rightarrow relevant
 N - not relevant

20
20

(2) Precision ≥ 0.5 (2)

$$MAP = \frac{1 + 1 + \frac{3}{4} + \frac{4}{8} + \frac{5}{10} + \frac{6}{20}}{6} = 0.675$$

(3) Precision ≥ 0.3 (3)
 R - Precision = $\frac{1}{2}$

④

① ס'ה פ'ס'ק ג' ה'ב'ט'ל'א :

פ'טל'הו כ'הו "א'ח'ל'ור מ'י'ד'ה" ו'ס'ל'ג' מ'ש'מ'ל'ה כ'י'ו'ן ש'כ'הו
 ג'צ'כ'פ' א'פ' מ'י'ל'ו'ר ע' מ'י'ל'ו'ר ע'ל'ו'ל ז'י'ב'ור פ'פ' מ'ק'ב'ו'ס א'ו'
 ק'ור'ס ג'מ'ק'ר'ה ש'ל'ו' ו'ל' "פ'פ'" ל'ג' ו'ס' מ'ש'מ'ל'ה ו'ג'ד'ר'ה מ'כ'ו'י'ה'
 א'פ' י'ש'ל' מ'י'ל'ו'ר נ'ר'ב'ע'ל'ה א'נ'ח'ו ע'ל'ו'ל'ס ל'ז'ל'ל ג'נ'י'ס מ'ז'ו'ח'י'ס
 ז'ב'ו'ה'ה פ'מ'ב'ט'ל' א'ה' ה'מ'י'ל'ו'ר : "ג'ל'" - א'י'ש'מ'י'ב'י'ג'י' ש'י'ס א'ו'
 מ'ש'מ'ל'ה ש'ו'נ'ה ל'ג'מ'ח'י' מ'א'ח'ל'ו'ר מ'י'ד'ה. א'ח'ל'ו'ר מ'י'ד'ה ל'ג' י'ב'ל'
 ז'ל'ו'ה ג'ע'ז'ס פ'פ' ש'ל' מ'ק'הו (ק'ור'ס, מ'ק'ב'ו'ס) ז'ב'ל'" - א'י'ש'מ'י'ב'י'ג'י'
 ל'ג' ו'א'ו'ר ש'ל' ע'ט'ל'ג'. כ'ל'ז'ה א'פ' י'ס מ'ס'מ'כ'י'ס פ'מ'י'ל'ו'ר ה'נ'ר'ב'ע'ל'ה
 מ'א'פ'י'צ'י'ס ג'ה'ס ג'ה'ק'י'ר א'ח'ר מ'ח'ו' פ'ר'ב'י'ס ה'ס א'ל'ו'ל'ס ל'ח'ל'ו'ר
 ו'ט'ל' ל'ח'ל'ו'ר ג'ב'י'ח' ו'ה' ה'מ'י'ל'ו'ר ה'מ'ח'ו'ה ש'נ'ש'ל'ו' ג'ש'א'ל'ו'ה.
 ל'ה נ'כ'ו'ן ג'ס' ל'ח'ל'ו'ל'ס ע'ק'מ'ס, ל'ו'י'ק' ל'ע'ב'ד'ו'ה. ו'ל' א'פ'
 ס'י'ב'ה ל'ר'ס'ת' ה'מ'ז'ב'ט'ל'ה :

[illegible][illegible]

(2-) יוחי לא, שהם את השאלות (על שם הרב ג' יחיא)
באר את השאלות הוחלט בהם, ונעזר באברהם.
שהם ככה את כל השאלות (כאשר לא כל השאלות) וכל
כמה שאלות יש חלק בין כל השאלות של השאלות
לשאלות המקוריות. שאלות נוספות, מספר השאלות

[illegible]

ХИМИКАТ

אנו מברכים את הורינו הבור

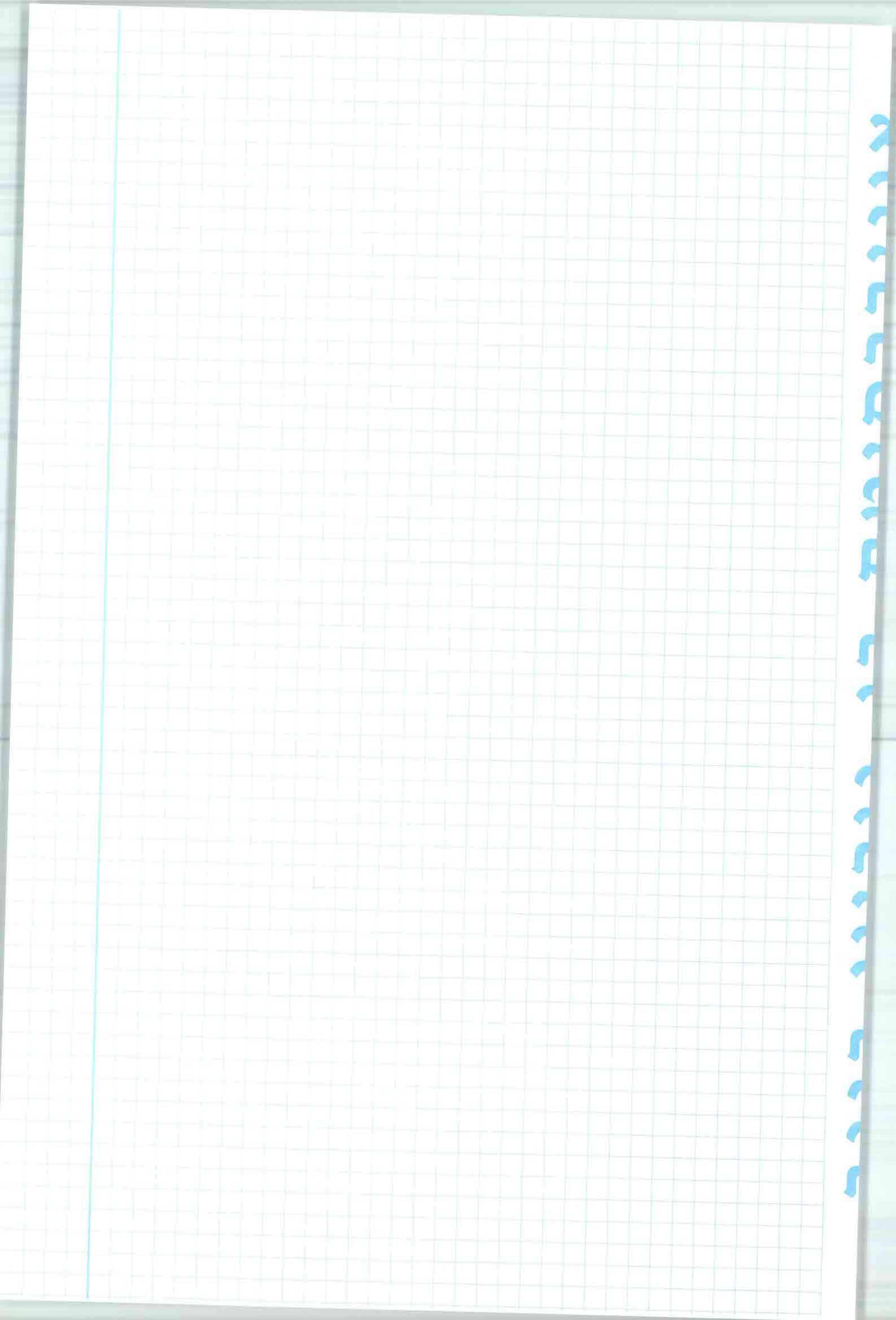
ALPHABETICALLY

אנו מרחיבים את גורמינו

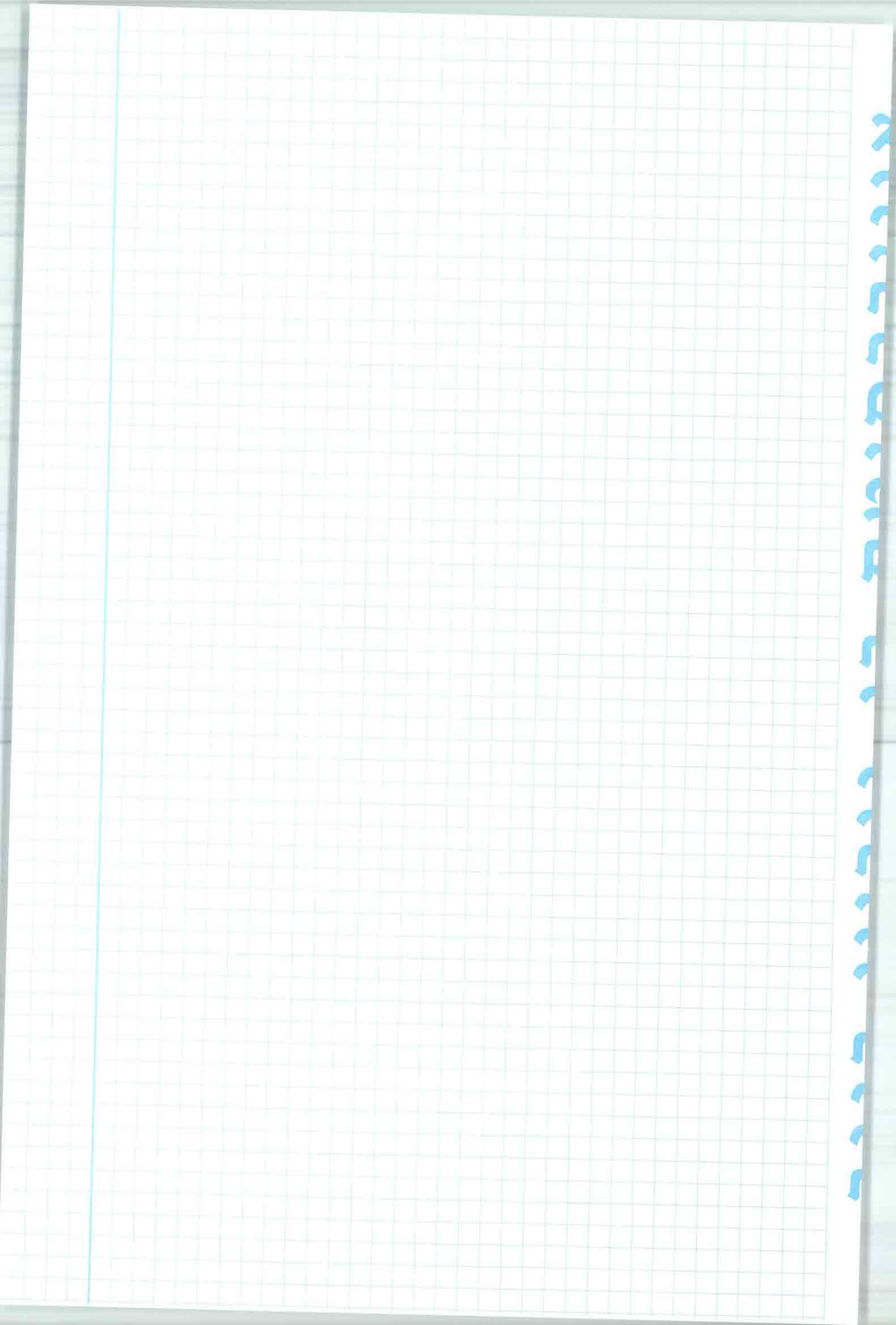
ХИТТОУОИ ТЛЕХИЛ ТЕТ

אנו מודים לך - רונדל

ХИТЛОУОТЛ-ХИЛЛЕРТ



NOTES ON THE HISTORY OF THE



NOTES ON THE HISTORY OF THE

WITNESS TO THE DEED