



23/9/08

תאריך:

אל: מזכירות הפקולטה למדעי ההנדסה

מאת: אריאל

שלום רב,

הנדון: דו"ח סיכום בחינה

לידיעת הסטודנטים - ממוצע קורס נקבע רק לאחר קבלת תוצאות בחינת מועד ב'

1. שם הקורס: אמצ"ן וקצ"ן / מסלול ניהול  
מספר הקורס: 372-1-4406 קורס: חובה / בחירה  
שם המרצה: ד"ר חיה לסר

2. הקורס ניתן בסמסטר: א' ב' ג' קיץ שנה"ל תשס"ח והשתתפו בו 4 סטודנטים

בבחינת מועד ב'

בבחינת מועד א'

מספר הנבחנים:	<u>27</u>	מספר הנבחנים:	
הציון הממוצע:	<u>73.1</u>	הציון הממוצע:	
סטיית תקן:	<u>13.7</u>	סטיית תקן:	
אחוז הנכשלים:	<u>8.3%</u>	אחוז הנכשלים:	

תאריך: 23/9/08 חתימת המרצה: אורי

חתימת רמ"ח: 23/9/08 תאריך: 23/9/08

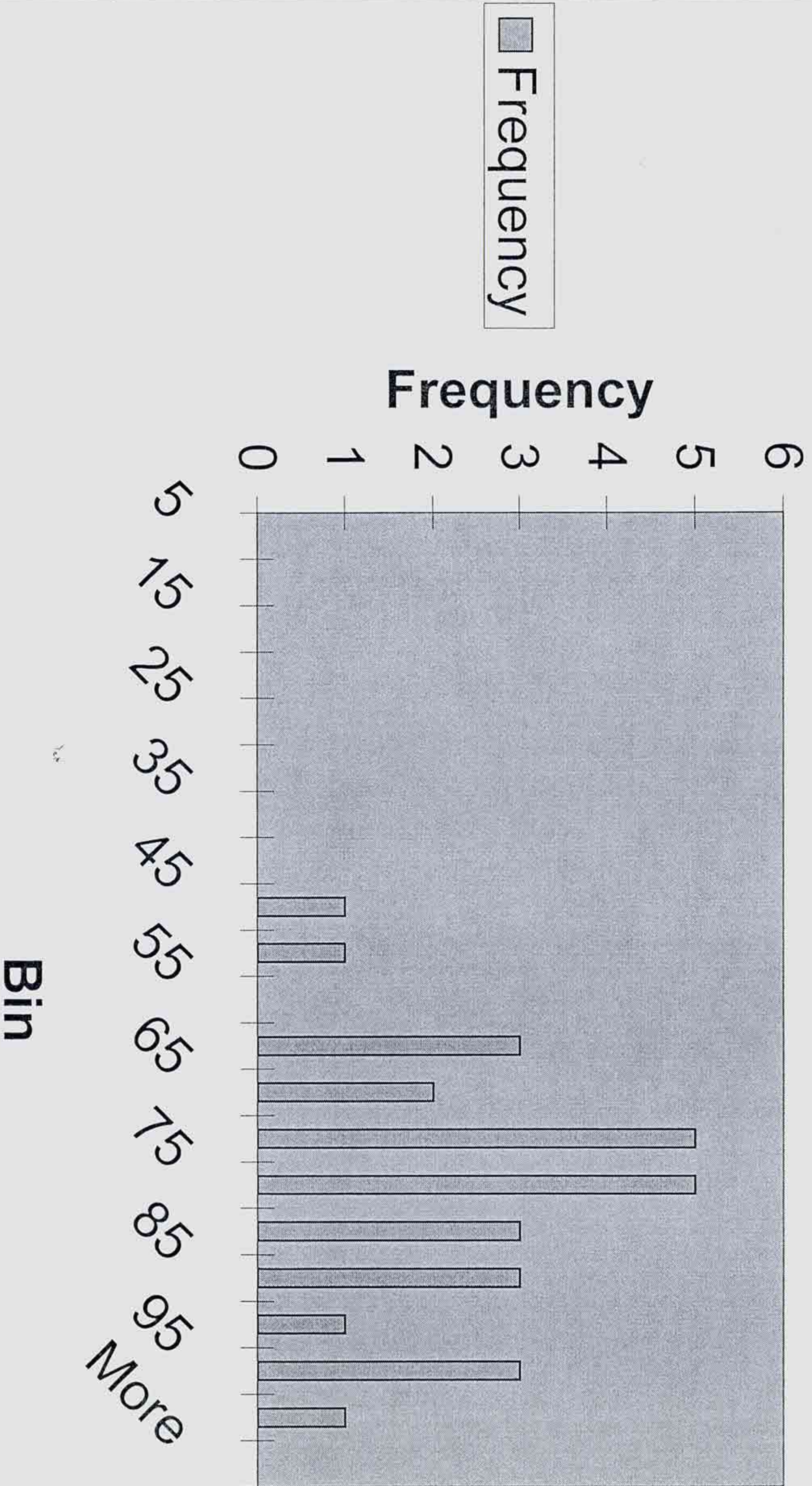
ממוצע קורס (יש לרשום רק לאחר תוצאות בחינת מועד ב') .....







# Histogram









בס"ד  
איחזור מידע וספריות דיגיטליות  
מועד ב', סמסטר ב' תשס"ח 16.09.08  
ד"ר ברכה שפירא, ארז שלום

חומר עזר מותר – לא מחשב נייד  
משך הבחינה : שעתיים וחצי

יש לענות על כל השאלות:

1. 20% נניח שאילתא Q1 שלה 10 מסמכים רלוונטים במאגר שבו 100 מסמכים. מנוע החזיר 6 מסמכים על פי הסדר הבא:

+1  
-2  
+3  
+4  
-5  
-6

כאשר + מסמן מסמך רלוונטי ו - מסמן לא רלוונטי.

א. 10% חשב: precision, recall, MAP, R-precision ו average precision של המנוע על פי השאילתא הנ"ל.

ב. 10% נניח שבמאגר 10,000 מסמכים (במקום 100), אבל לשאילתא Q1 יש בדיוק אותם מסמכים רלוונטים במאגר. גם כשהמנוע הפעיל את Q1 על המאגר הגדול יותר, חזרו בדיוק אותם מסמכים. באופן אינטואיטיבי ביצועי המנוע טובים יותר על המאגר הגדול כי היו לו יותר מסמכים לא רלוונטים לאתר מתוכם את הרלוונטים. הסבר מדוע המדדים שחישבת ב-א לא מצליחים לזהות את הביצועים הטובים יותר של המנוע על המאגר הגדול, והצע מדד חדש שיבטא את ההבדל בביצועים.

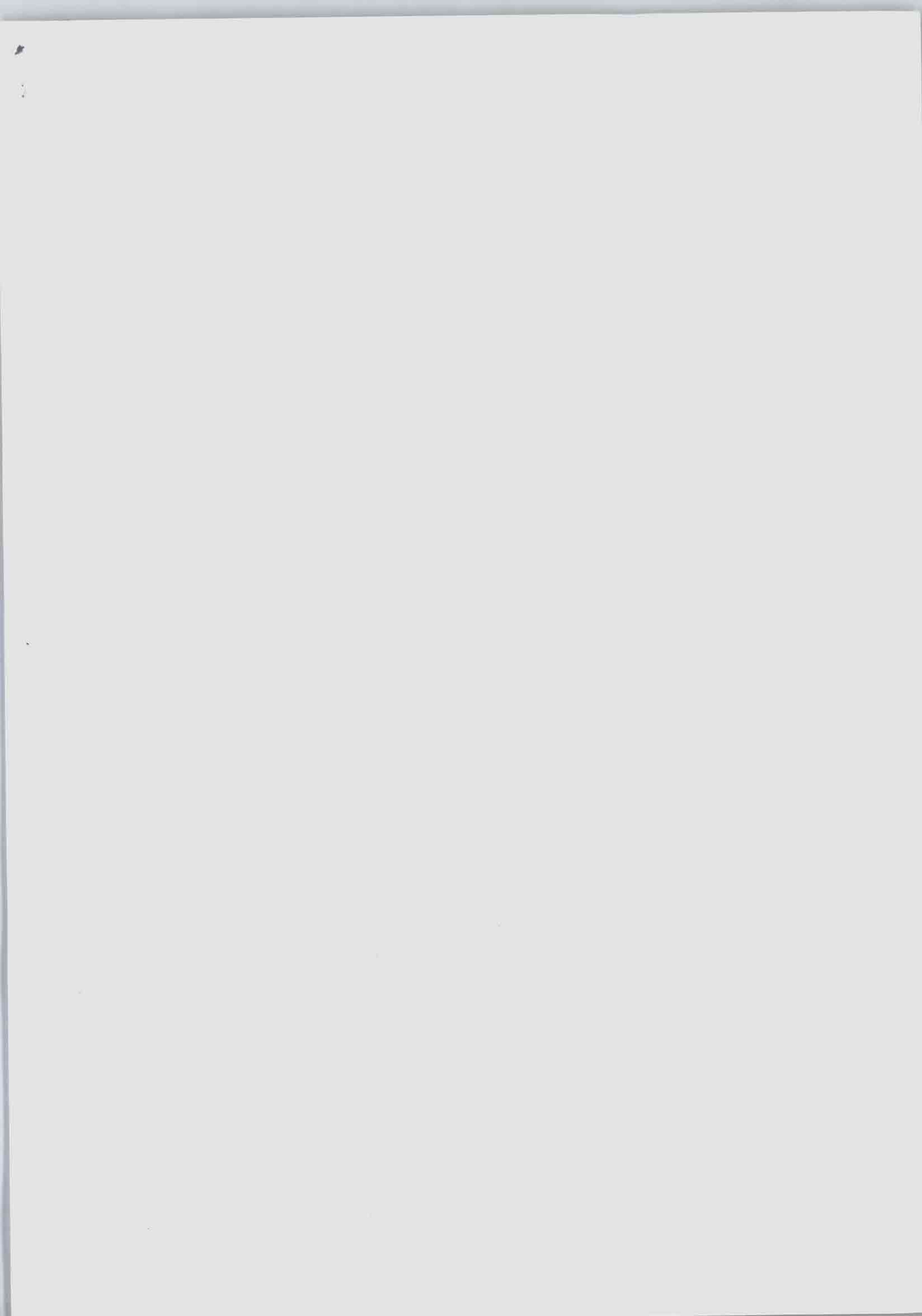
2. 25% הצע אלגוריתם לסיווג של spam למערכת של דואר אלקטרוני. כלומר, אלגוריתם שבודק על כל דואר שמגיע האם הוא ספאם או לא. האלגוריתם צריך להתבסס על שימוש במודל הווקטורי לצורך מידול הווקטור של הדואר ה"חוקי" ושימוש בנוסחת Rocchio. הסבר על סמך מה ייבנה הפרופיל ה"חוקי", איך הוא יתעדכן ועל פי מה תתקבל ההחלטה אם לסנן כל הודעת דואר.

3. 20%

א. 10% הנח שמשתמשים בשיטה הבאה להעריך גודל יחסי של מנועי חיפוש A ו B. המנועים מגרילים דפים באופן רנדומי מהאינדקסים שלהם ובודקים האם הדפים שהגרילו מאונדקסים במנוע השני. נמצא ש 30% מהדפים ש A הגריל נמצאו אצל B, 50% מהדפים שהגריל B נמצאו אצל A. מהו היחס בין מספר הדפים ש A מאנדקס לעומת B.

ב. 10% איך אפשר להשתמש בידע על מבנה ה WEB (מבנה עניבת הפרפר) באופן מעשי. הראה דוגמא ספציפית.

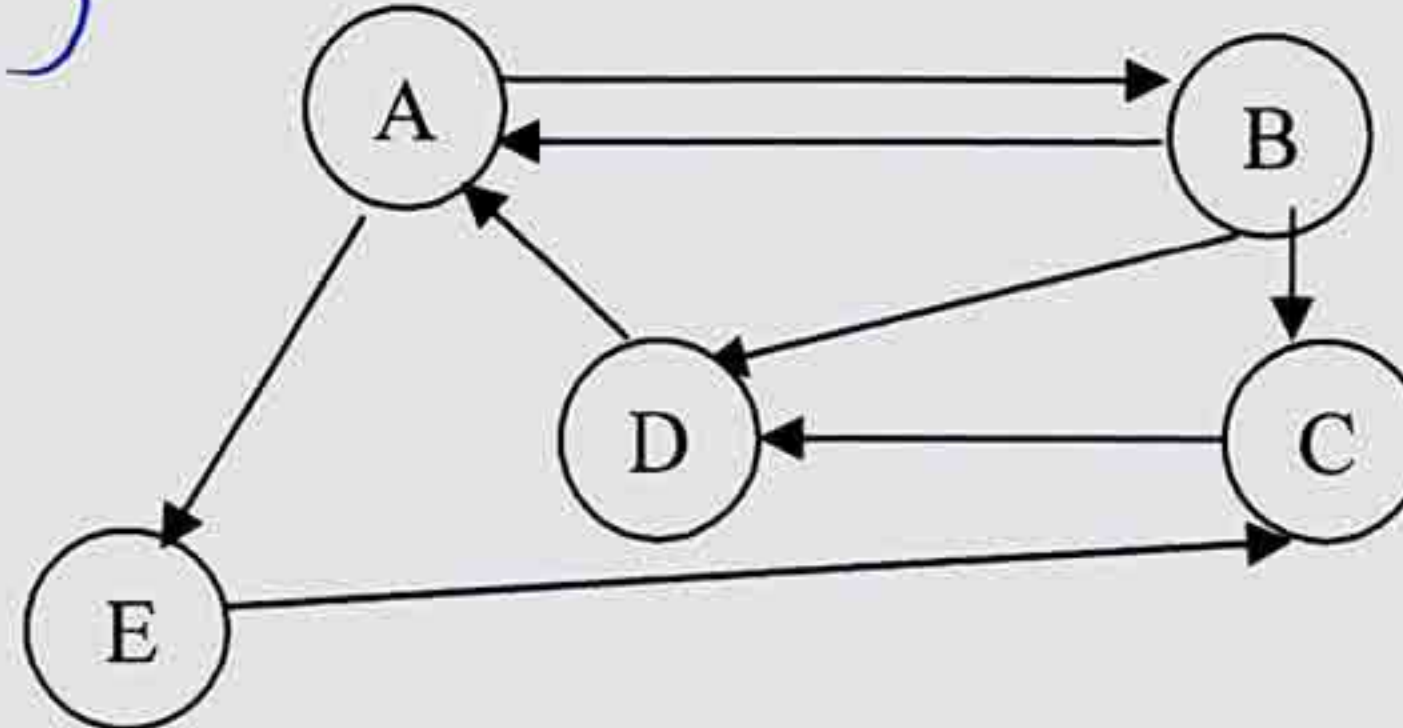




4. 15% בדף מסויים שידוע שהוא מאונדקס על ידי מנוע חיפוש מסויים, אפשר למצוא אוסף של מילים שאם נפעיל אותם כשאילתא למנוע הדף לא יוחזר. תאר לפחות שתי סיבות למצב כזה. (תשובה כללית כמו שהמנוע לא טוב, לא מתקבלת, צריך להסביר על דרך אפשרית של פעולת המנוע שיכולה לגרום לזה)

Heavy Ranking

5. 20% נתן הגרף הבא של Web



- א. 10%. מיהו הצומת בעל ה- Pagerank גבוה ביותר, מיהו הצומת בעל ה- pagerank הנמוך ביותר. הסבר את תשובתך (אין צורך לבצע חישוב)
- ב. 10%. האם באופן כללי חישוב של Authority על פי אלגוריתם HITS מקביל לחישוב pagerank? (בהנחה ששניהם מופעלים על כל הרשת). הסבר אם תוצאת החישוב דומה, או את ההבדל (תשובה שמציגה את הנוסחאות השונות בלי הסבר לא מתקבלת, יש להסביר את המשמעות והסיבה להבדל או הדמיון).

בהצלחה

ברכה וארז





372-1-4406

אחרי מניח אספה דוגמא  
 צד משקל - מולד

$\frac{3}{6} = pr$  (1)

$\frac{3}{10} = result$

$(\frac{1}{1} + \frac{2}{3} + \frac{3}{4}) / 3 = avg$

$(\frac{1}{1} + \frac{2}{3} + \frac{3}{4} + 0 \dots) / 11 = Avg$

האסר משה רק כי לא יאזכר מה לאונטה בתור החצי בסו  
 משיכה. אם נניח שהמקור לא החצי יאזכר לאונטה אז זה  $\frac{3}{6}$

(2) האסר fallat או מצבים אחרים יחסים למחשבים בקול המוקד.

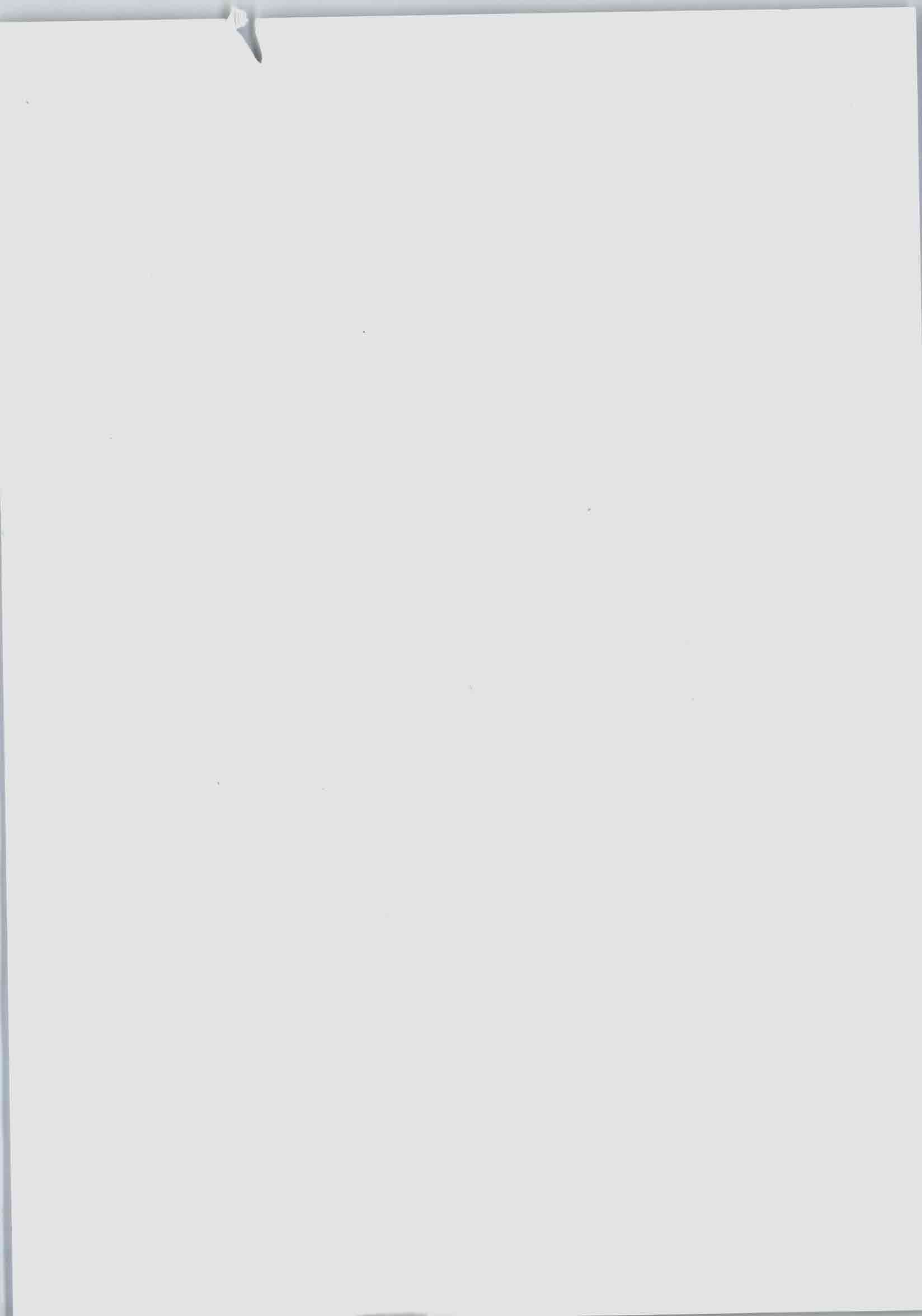
(כיוון:  $\frac{1 - \text{שטח}}{\text{שטח}}$ )  
 - בדונטל סאטור ע"י set & הוצאת שמות למעלה

אשר לא יאזכר ע"י הוצאה אם הוא spam או לא. המוקד  
 הואשור למעלה אוקטו (פירסר האשור) אפוא הוצר  
 לעצמה ~~הוצאה~~ הוצאה אוקטו הוצר או האשור או  
 feedback יאזכרם Rochio.

היון כי ~~האשור~~ למעלה אוקטו - אשור דג"ן (10)  
 בין האשור דג"ן הפירסר א האשור. (100)

הוא הוצר למעלה אשור מוקד מוקד feedback פירסר  
 הוא spam. אפי ה-feedback אוקטו אוקטו הפירסר  
 אפי נאמר Rochio.







$$A = \frac{5}{3}B \quad (3)$$

② - אלו הם מונחים שציינת ביום רבין עצים ואלן ה crawler

צין אחר Loops

- אלו כיוון ה crawler איננו ממוקד קולר הנחש, כפי ש' א' איננו ממוקד (אלא ב out)

- לא עברו לציין איננו ב - out וכן - הגלגל

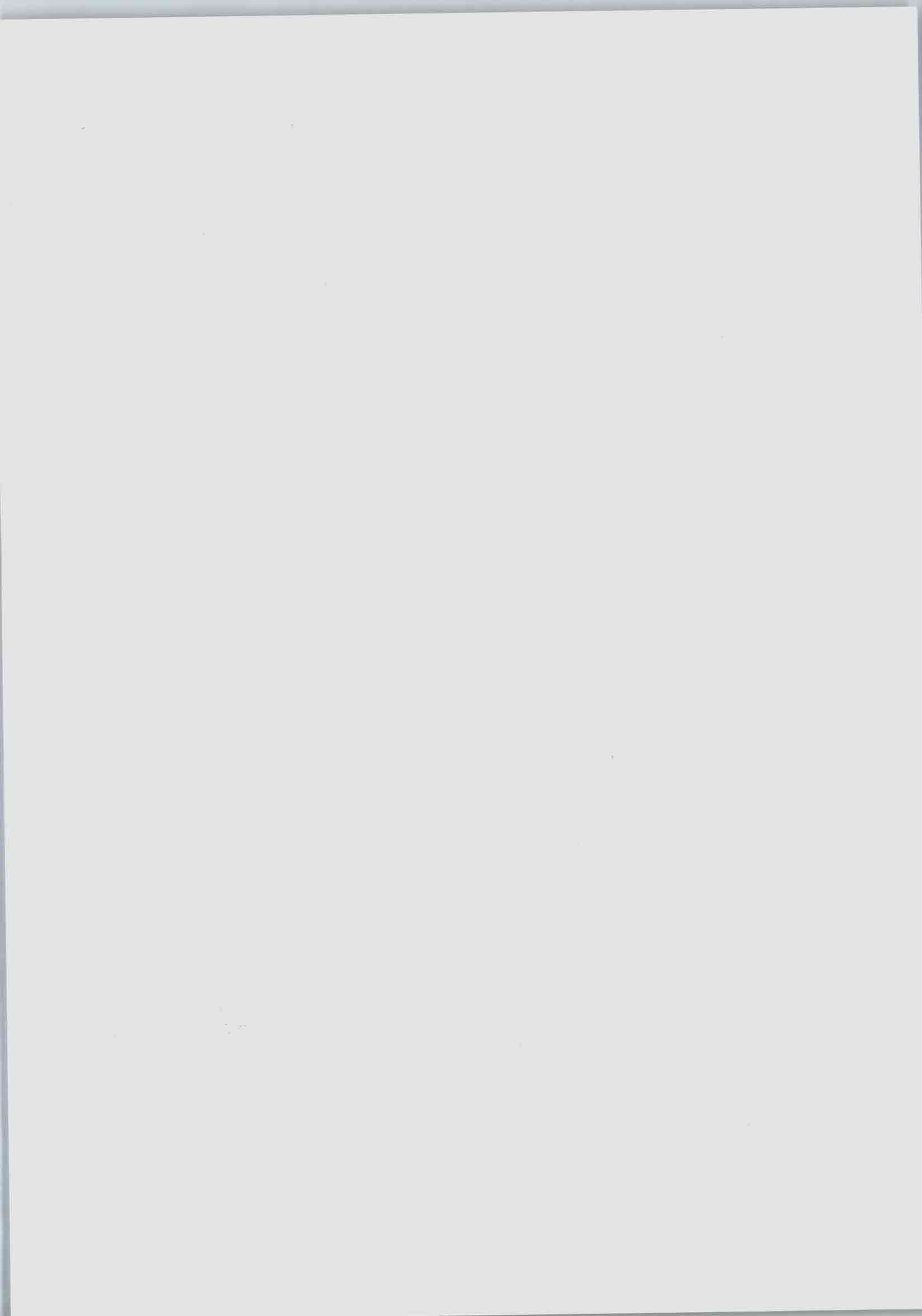
נראה שהם לא יודעים על הדמיון באיזה אזור בלבד (ין שיש מאגרים בהבון)

④ - אלו מונחים מאגרים ~~ה~~ ין חק א הצי.

- לא הצי מאגרים לא מאגרים אלא ין Anchor

- לא מונחים שיהיו stop-words







$$(1) PR(A) = PR(B)/3 + PR(D)$$

(1) (5)

$$(2) PR(A) > PR(D)$$

$$(3) PR(E) = PR(A)/2$$

$$(4) PR(B) = PR(A)/2 \Rightarrow (5) PR(E) = PR(B)$$

$$(6) PR(C) = PR(B)/3 + PR(E) \Rightarrow (7) PR(C) > PR(E)$$

$$(8) 5, 7 \Rightarrow PR(C) > PR(B)$$

$$(9) PR(D) = PR(B)/3 + PR(C) \Rightarrow (10) PR(D) > PR(C)$$

$$(11) 2, 10 \Rightarrow PR(A) > PR(C)$$

$$A > C (11)$$

$$A > B (8, 11)$$

$$A > D (2)$$

$$A > E (E=B)$$

$$PR \text{ of } E, B$$

$$A - \text{העובד ב'אמר'}$$

(2) לחלוקה

Authority - אמרנו כי הם א הוציאו שני דברים וכל אחד מהם

הוציאו את המידע הזה.

PR - אמרנו כי הם א הוציאו שני דברים וכל אחד מהם







יון בנגב

קבוצה: סמל המקצוע: 7-12

372 הנדסת מערכות מידע

המחלקה:

מחזור מידע וספריות דיגיטלי 0-372-1-4406

המקצוע:

ד"ר ב. שפירא

שם המורה:

85062

מועד: 21 באוניברסיטת בן-גוריון

2008-1  
18-19-20

סמסטר:

מיון

דף: 1

שעה: 09:00

16.09.2008

תאריך המבחן:

חדר-

מס' זהות	מס' נבחן	מס' סידורי	ציון סופי	ציון	ציון
21768304	5303	1/1		56	
21904008	9554	2/2		71	
26626002		3/3			
32903080		4/4			
36249688	5399	5/5		79	
36307494	9560	6/6		64	
37714391		7/7			
37723954		8/8			
37768942		9/9			
38068086		10/10			
38763728		11/11			
38815254		12/12			
39012943		13/13			
39015110		14/14			
39245014		15/15			
39351549	9553	16/16		50	
39593496		17/17			
39610589	9561	18/18		73	
39735675	9555	19/19		67	
3979372	5377	20/20		80	
סה"כ					

חתימת המורה האחראי

תאריך:

תאריך פרסום:

חתימת ראש המחלקה

המחלקה (3) לוח מודעות 4 מורה

יון בנגב  
הנדסה

ת







יו"ן בנגב

המחלקה: 372 הנדסת מערכות מידע סמל המקצוע: קבוצה 7-12

המקצוע: אחריות מידע וספריות דיגיטליות 0-372-1-4406

שם המורה: ד"ר ב. שפירא 850621

מסמט: 2008-1 18-19-20 מועד: ב' אונ' ברסינת בן-ג 21

תאריך המבחן: 16.09.2008 שעה: 09:00 דף: 2

מס' זהות	מס' נבחן	מס' סידורי	ציון סופי	ציון	ציון
39867692		21/1			
39869474		22/2			
39945837		23/3			
39974316	9556	24/4		89	
39981923		25/5			
39986294	5383	26/6		82	
39999917		27/7			
40186538		28/8			
40295149		29/9			
40297954	5302	30/10		44	
40373219	5044	31/11		85	
40546293		32/12			
40806499		33/13			
40858961		34/14			
43389394		35/15			
43397405		36/16			
43421734		37/17			
43463389		38/18			
43490077		39/19			
46266	5304	40/20			

סה"כ

חתימת המורה האחראי

תאריך:

תאריך פרסום:

חתימת ראש המחלקה

מחלקה: 3. לוח מודעות 4. מורה

יו"ן בנגב  
התנדבות  
יכרות

24

למה

23/9/08

23/9







ייון בנגב

המחלקה: 372 הו דסת מערכות מידע סמל המקצוע: קבוצה 7-12

המקצוע: אחזור מידע וספריות דיגיטלין 0-372-1-4406

שם המורה: ד"ר ב. שפירא 85062

מידע: 1-6 סמסטר: 2008-ב מועד: ב' אונ' ברסינת בן-ג 21

תאריך המבחן: 16.09.2008 שעה: 09:00 דף: 3

מס' זהות	מס' נבחן	מס' סידורי	ציון סופי	ציון	ציון
60139011	5376	41/1		68	
60462843	5398	42/2		92	
60463445		43/3			
60631603		44/4			
60719903		45/5			
61114716		46/6			
61287074		47/7			
61678371	5300	48/8		85	
62851332		49/9			
62883350	5384	50/10		94	
64976053	5301	51/11		78	
64978307		52/12			
65676397	9557	53/13		62	
65911240		54/14			
65949265	9558	55/15		68	
66127937	5397	56/16		58	
66132820		57/17			
66149188		58/18			
66196072		59/19			
301209268		60/20			

סה"כ

חתימת המורה האחראי

תאריך:

תאריך פרסום:

חתימת ראש המחלקה

3. לוח מודעות 4. מורה

גב  
ה

23/9/08

23/9







המחלקה: 372 הנדסת מערכות מידע סמל המקצוע: קבוצה:  
7-12

המקצוע: אחזרז מידע וספריות דיגיטלי 0-372-1-4406

שם המורה: ד"ר ב. שפיידר 85062

מועד: ב' אלול תשס"ח 21-10-2008 סמסטר: 1-2008

תאריך המבחן: 18.09.2008 שעה: 09:00 דף: 4

## סה"כ

תאריך פרסום:

23/9/07  
1290

17/3/9

המחלקה 3. לוח מודעות 4. מורה

דרייזן  
ההנדסה  
עבודת

24



