



תאריך: 05/02/2011

אל: מזכירות הפקולטה למדעי ההנדסה

מאת: בב"ס ספי

שלום רב,

הנדון: דו"ח סיכום בחינה

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

1. שם הקורס: אמצעי מידע וספריות דיגיטליות

מספר הקורס: 372-1-4406 קורס: חובה / בחירה

שם המרצה: בב"ס ספי

2. הקורס ניתן בסמסטר: (א') ב' ג' קיץ שנה"ל תשע"א והשתתפו בו 118 סטודנטים

בבחינת מועד ב'

בבחינת מועד א'

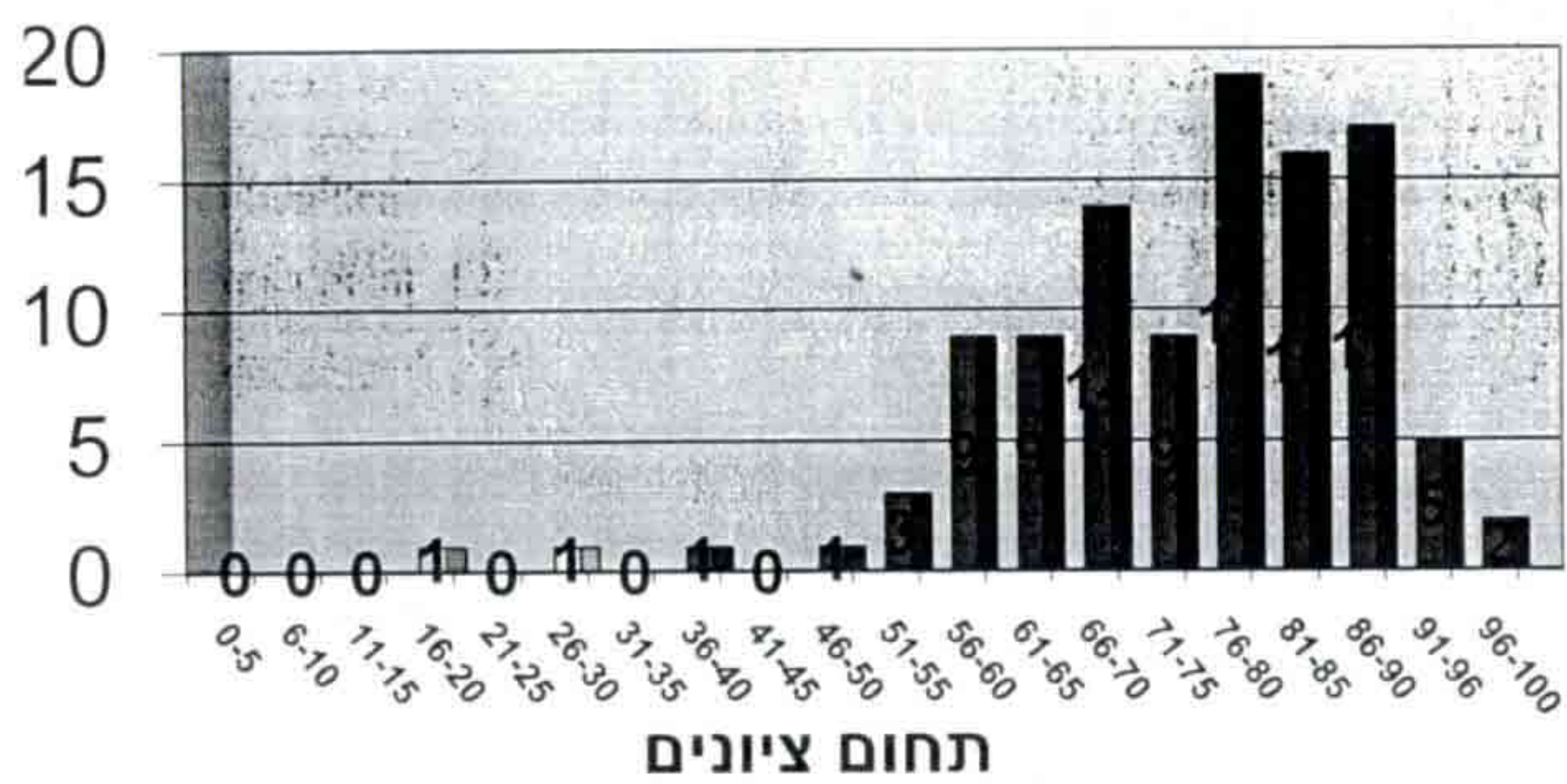
מספר הנבחנים:	<u>107</u>	מספר הנבחנים:
הציון הממוצע:	<u>74.4</u>	הציון הממוצע:
סטיית תקן:	<u>14</u>	סטיית תקן:
אחוז הנכשלים:	<u>7%</u>	אחוז הנכשלים:

חתימת המרצה: בב"ס ספי תאריך: 05/02/2011

חתימת רמ"ח: [חתימה] תאריך: 7/2/11

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

התפלגות ציונים





איחזור מידע תשע"א – 372.1.4406

סמסטר חורף מועד א' 31.01.11

ד"ר ברכה שפירא, אורלי מורנו

מספר נבחן _____

משך המבחן : שעתיים וחצי

חומר עזר: מותר (לא מחשב נייד) – יש להחזיר את הטופס

חלק א.

סמן תשובה אחת נכונה –יש לסמן תשובות על טופס המבחן.:

1. 4% מנוע חיפוש מחזיר לכל שאלתא 10 מסמכים (מתוכם, חלקם רלוונטיים). הוחלט לנסות לשפר את ביצועי המנוע על ידי הגדלת מספר המסמכים המוחזרים לכל שאלתא מ 10 ל 1000, כלומר, יוחזרו עוד 990 מסמכים על 10 המוחזרים.

אם אף מסמך נוסף לא זוהה כרלוונטי בקבוצת המסמכים הנוספת המוחזרת אז :

- א. ה- Precision ו ה Recall של המנוע יישארו אותו דבר גם לאחר הוספת המסמכים לתוצאה
- ב. לאחר ההוספה, ה precision יקטן ו ה Recall יישאר אותו דבר
- ג. ה Precision וה Recall יקטנו לאחר ההוספה
- ד. ה r-precision יקטן
- ה. אף תשובה לא נכונה

2. 4% (בהמשך ל-1). אם לפחות מסמך אחד נוסף זוהה כרלבנטי בקבוצת המסמכים הנוספת שהוחזרה, אז:

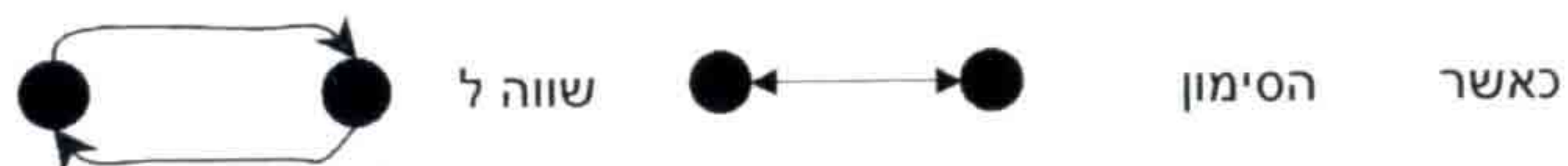
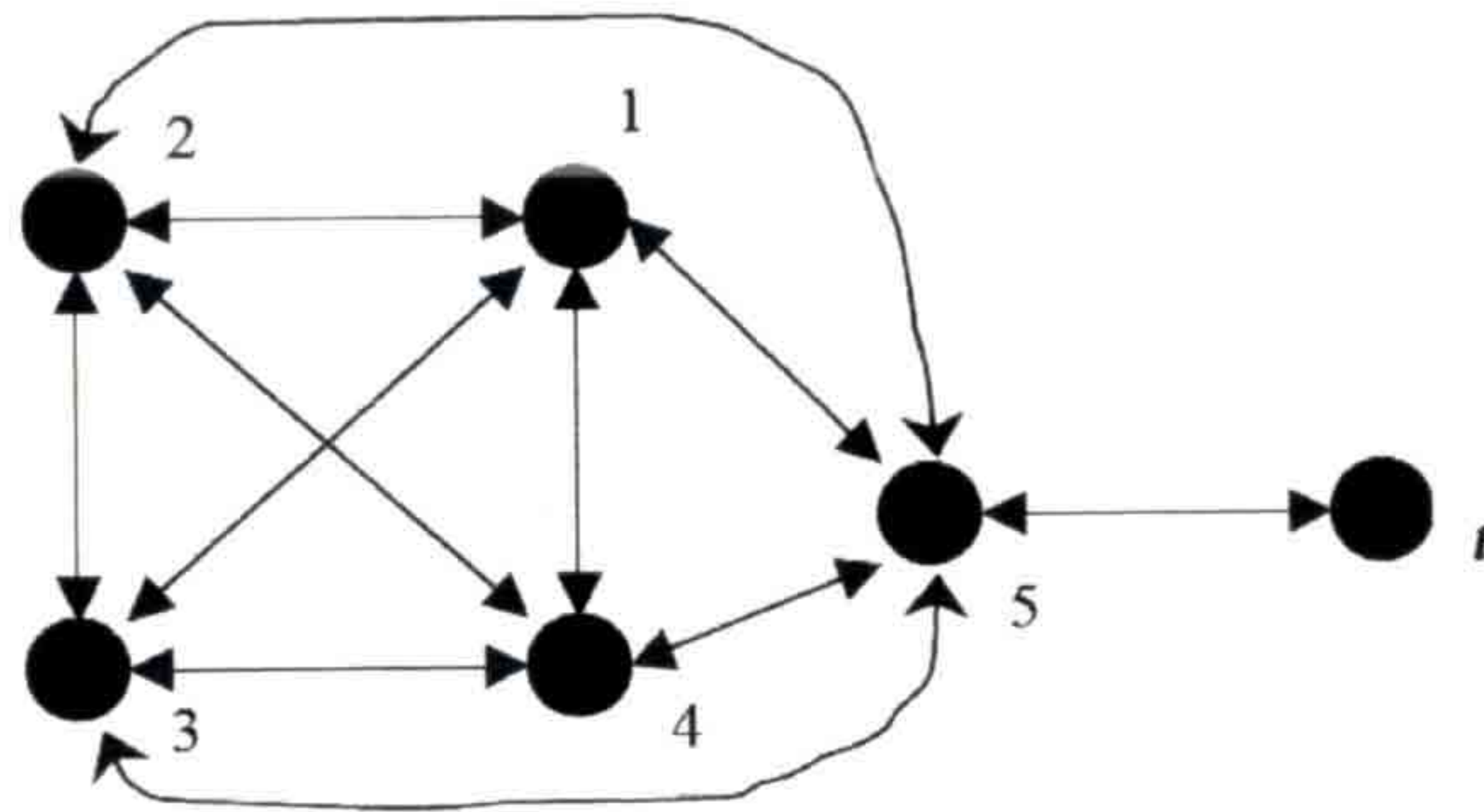
- א. ה precision לאחר ההוספה יהיה בהכרח נמוך יותר
- ב. ה precision לאחר ההוספה יהיה בהכרח גבוה יותר
- ג. ה Recall לאחר ההוספה יהיה בהכרח גבוה יותר
- ד. ב-ו-ג נכונים
- ה. ה precision ב – 10 מסמכים יגדל

3. 4% Interpolated average precision הוא מדד יעיל ל-

- א. השוואה בין מנועים ששונים רק בסדר המסמכים שמוחזרים
- ב. השוואה בין מנועים שמתחשב גם ב Recall
- ג. השוואה רק בין ה precision של המנועים
- ד. ציור של גרף precision-recall על פני הרבה שאלות
- ה. ב+ד נכונים



4. 8% נתונה הרשת הבאה :



לרשת שכללה רק את הצמתים 1-5, נוספה הצומת t, כיצד ישתנו ערכי ה pagerank של הצמתים ברשת בעקבות ההוספה:

- כל ערכי ה pagerank של כל הצמתים ישארו שווים לערכם לפני ההוספה, גם ערך צומת t יהיה שווה לערכי צמתים.
- הערך של הצמתים 1-4 יקטנו במידה שווה.
- הערך של צומת 5 יגדל ויהיה הגבוה ביותר ברשת.
- הערך של צומת t יהיה הקטן ביותר ברשת.
- הערך של צומת 5 יקטן.
- ה+ב נכונים.
- ב+ג+ד נכונים.
- רק ב+ג נכונים.

5. 16% ציין נכון או לא נכון ליד כל אחד מהמשפטים הבאים:

- 4% במנוע חיפוש באינטרנט F-measure הוא מדד חשוב.
- 4% נניח מאגר מסמכים המכיל את המסמך השגוי הבא: "שלום טמבל מה העניינים ואיך אתה מרגיש" - נניח מנוע הפועל לפי המודל הווקטורי. נניח שהטעות (מופע של המילה "טמבל" לא רלוונטית למסמך) התגלתה. תיקון הטעות ישפיע רק על המסמך השגוי.
- 4% שאילתא הכוללת מילה שיש לה מילים נרדפות רבות בהכרח תסבול מ precision נמוך.
- 4% תור של Crawler של מנוע המיועד לבניית מאגר בנושא כלשהו יכלול רק דפים הרלוונטים לנושא המאגר.

חלק ב.

ענה על השאלות הבאות:

6. 7% נתונה טבלת הדירוגים הבאה של 5 משתמשים (user1..user5) לחמישה items (A..E). כמו כן נתונה הקרבה ($W_{5,i}$) בין כל אחד מה-users ל user5 (המוגדר כ Active user). יש לנבא את הרלוונטיות של D item ל user 5 על פי שני השכנים הקרובים אליו ביותר (יש לחשב את $(P_{5,D})$).

Item	User1	User2	User3	User4	User5 (active user)
A	10	5	9		9
B	6	9		5	5
C	2	7	3		1
D	4	8	3	3	
E	8	1	9	2	
דמיון ל user5 $W_{5,i}$	1	-0.5	0.9	0.7	

7. 15% הצע מדד Discount Cumulative Gain (DCG) קבוצתי, שמפחית ציון של מסמכים בקבוצות של 5. כלומר הפחתה מסוימת ל- 5 המסמכים הראשונים, הפחתה מוגדלת ל-5 מסמכים השניים וכו'. במסמכים שבתוך כל קבוצה ההפחתה שווה. כתוב נוסחה מדויקת ל DCG קבוצתי לדרגה K כלשהי, הנוסחה צריכה לכלול תמיכה ב- K שאינו בכפולות של 5. יש להסביר בקצרה את ההיגיון של הנוסחה.

8. 42% במאגר של 25000 מסמכים צירפו לכל מסמך רשימה מדורגת (לא ריקה) הכוללת 1-10 מילים לכל מסמך של מילות מפתח מתוך מאגר של 100 מילות מפתח הרלוונטיות למאגר (המילים לא בהכרח מופיעות במסמך אליו הן מצורפות). סדר המילים המצורפות למסמך מעידה על חשיבותן למסמך (המילה הראשונה חשובה ביותר, השנייה פחות מהראשונה, השלישית פחות מהשנייה וכו'). שאילתא במנוע שיועד למאגר היא רשימה של מילות מפתח מתוך 100 המילים (ללא חשיבות לסדר המילים בשאילתא), אין משקל למילים בשאילתא.
- פונקציית הדירוג של מסמך לשאילתא צריכה :
- עבור שאילתא של מילה אחת להחזיר מסמכים שברשימה של מילות המפתח שלהם מופיעה מילת המפתח של השאילתא ולדרג את המסמכים המוחזרים לפי המיקום של המילה ברשימת מילות המפתח של המסמך.
- עבור שאילתא של יותר ממילה אחת, פונקציית הדירוג צריכה להתחשב במספר המילים המשותפות בין מילות המפתח של המסמך ומילות השאילתא וכן בסדר של מילות המפתח במסמך.
- (לדוגמא, מילות המפתח של מסמך A כוללות שתי מילים של השאילתא- מילה אחת של השאילתא היא ראשונה במילות המפתח של המסמך, והמילה השניה של השאילתא היא השניה במילות המפתח של מסמך A,
- מילות המפתח של מסמך B גם כוללות שתי מילים מהשאילתא- מילה אחת של השאילתא נמצאת במקום השלישי במילות המפתח של המסמך, והשניה במקום ה-5 במסמך. A ידורג יותר גבוה מ B.)
- א. 10% הצע מודל ווקטורי (עם 100 ממדים) שיתמוך בפונקציית הדירוג הנ"ל: תאר מה יכלול ווקטור של מסמך. תאר בקיצור כיצד נקבע המשקל של מילת מפתח בווקטור של מסמך (כלומר, מהי פונקציית המשקול של מילה במסמך). אין צורך לחשוב על מודל חסכני בזכרון.
- ב. 12% כתוב פונקציות דרוג מדוייקות שיבטאו את דרישת הדירוג לעיל. פונקציה אחת לשאילתות של מילה אחת, ופונקציה שניה לשאילתות עם יותר ממילה 1.
- ג. 14% כיצד ניתן לשלב את המודל שהצעת ב-א עם מודל ווקטורי סטנדרטי – כלומר מסמך ייוצג גם על ידי מילות מפתח מהרשימה וגם על ידי המילים המופיעות במסמך ושאילתא תוכל לכלול גם מילים מהמסמך וגם מילות מפתח.
- 6% כיצד יראה ווקטור של מסמך במודל המשולב
- 8% כיצד יחושב הדמיון בין שאילתא ומסמך (נוסחא מדויקת)
- ד. 6% ציין שיפור אפשרי של ביצועי מנוע המבוסס על המודל המוצע (כלומר, מדוע כדאי להשתמש במודל כזה, מה הוא יכול לשפר).

המחלקה: 372 הנדסת מערכות מידע סמל המקצוע: קבוצה: 1

המקצוע: אחזור מידע וספריות דיגיטליות 1. 4406 0-372

שם המורה: ד"ר ב. שפירא 125

סמסטר: א-2011 מועד: א' אוניברסיטת בן-גור

18-19-20

תאריך הסבת: 31.01.2011 שעה: 13:3 דף: 1

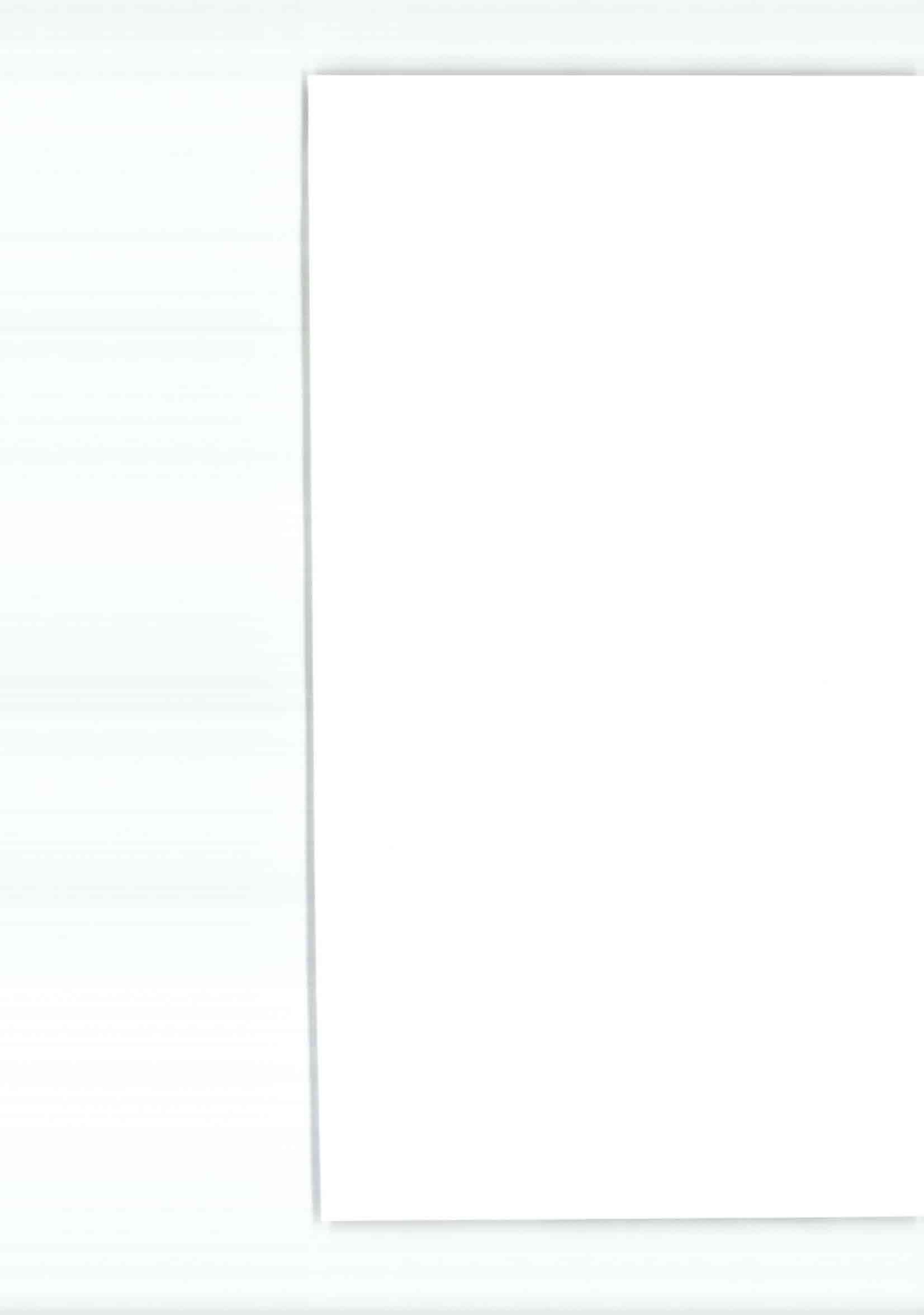
מס' זהות	מס' נבחן	מס' סידור	ציון סופי	ציון	ציון
43176676	✓	1/1		83	חד מחלק
66352311	✓	2/2		56	חד מחלק
66602004		3/3			חד מחלק
66602673	✓	4/4		90	חד מחלק
66629445	✓	5/5		82	חד מחלק
66651563		6/6			חד מחלק
66653874	✓	7/7		86	חד מחלק
200006468	✓	8/8		90	חד מחלק
200023950	✓	9/9		86	חד מחלק
200255511	✓	10/10		85	חד מחלק
300009602	✓	11/11		67	חד מחלק
300027760	✓	12/12		77	חד מחלק
300304128	✓	13/13		78	חד מחלק
301047916	✓	14/14		80	חד מחלק
301122115	✓	15/15		68	חד מחלק
301803235	✓	16/16		67	חד מחלק
302668900	✓	17/17		87	חד מחלק
302984471	✓	18/18		78	חד מחלק
302998646	✓	19/19		85	חד מחלק
304320054	✓	20/20		68	חד מחלק

סה"כ

בגב
דסה

13:3

מסמך הסבת



סה"כ



המחלקה: 372 הנדסת מערכות מידע סמל המקצוע: קבוצה 1: 7-12

המקצוע: אחזור מידע וספריות דיגיטליות 1. 4406 0-372

שם המורה: ד"ר ב. שפירא 125

סמסטר: א-2011 מועד: א' אוניברסיטת בן-גוריון 18-19-20

תאריך המבחן: 31. 01. 2011 שעה: 13:30 דף: 1

מס' זהות	מס' נבחן	מס' סידורי	ציון סופי	ציון /	ציון
306329228	✓	28/1		96	חד מחלק
306516188		29/2			חד מחלק
306696873	✓	30/3		52	חד מחלק
307028548	✓	31/4		86	חד מחלק
307105601	✓	32/5		86	חד מחלק
307498782	✓	33/6		95	חד מחלק
307699397	✓	34/7		74	חד מחלק
308663251	✓	35/8		56	חד מחלק
308734318	✓	36/9		71	חד מחלק
308868421	✓	37/10		16	חד מחלק
309253607	✓	38/11		90	חד מחלק
310316328	✓	39/12		74	חד מחלק
311693303	✓	40/13		66	חד מחלק
311785562	✓	41/14		56	חד מחלק
311979272	✓	42/15		80	חד מחלק
312892631	✓	43/16		79	חד מחלק
317058188	✓	44/17		89	חד מחלק
317263069	✓	45/18		68	חד מחלק
317385094	✓	46/19		81	חד מחלק
319381984	✓	47/20		88	חד מחלק
סה"כ					

912

סמל המקצוע: קבוצה 1:

7-12

אחזור מידע וספריות דיגיטלי 0-372.1.4406

ד"ר ב. שפירא

מועד: א' און יברסיטת בן-גור

2011-N

18-19-20

27

שנה : 3 : 13

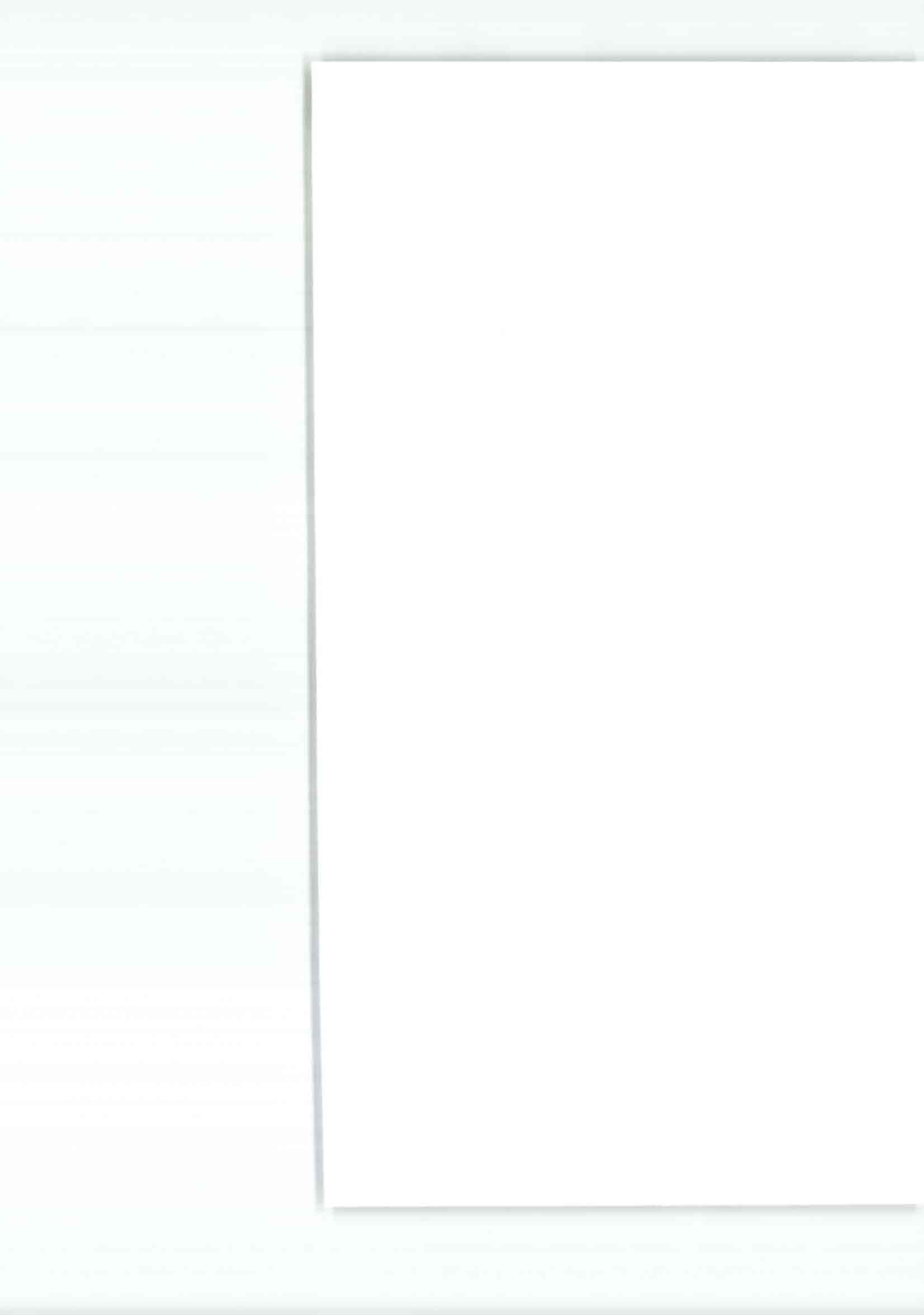
31. 01. 2011

204-777,

[illegible]

סה"כ

912



סמל המקצוע

המחלקה

7-12

הסקצו

שם המורה:

125

7510

סמסג

35

21

18-19-20

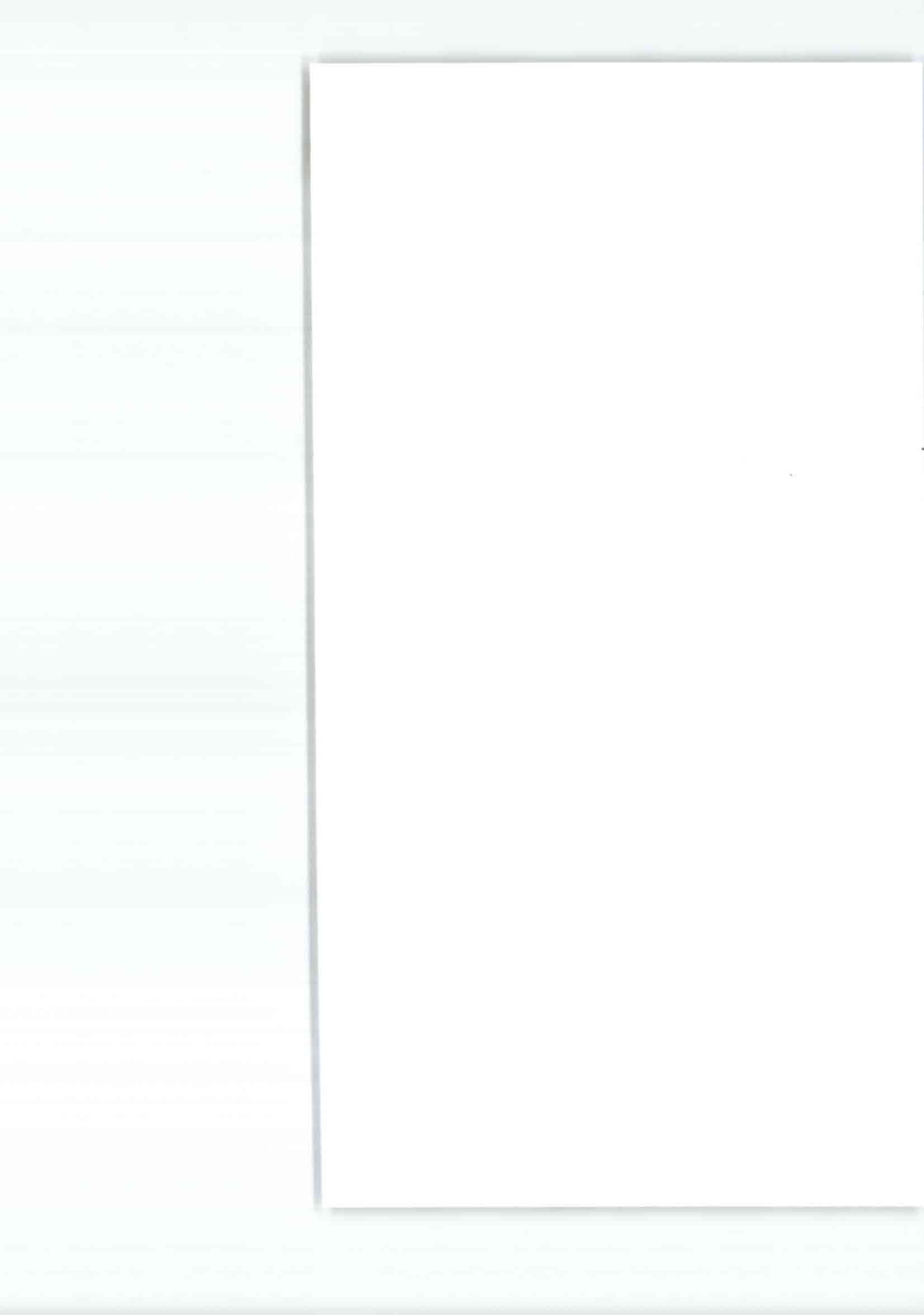
31. 01. 2011

204-77П,

177

בגד
ז'סיה





111

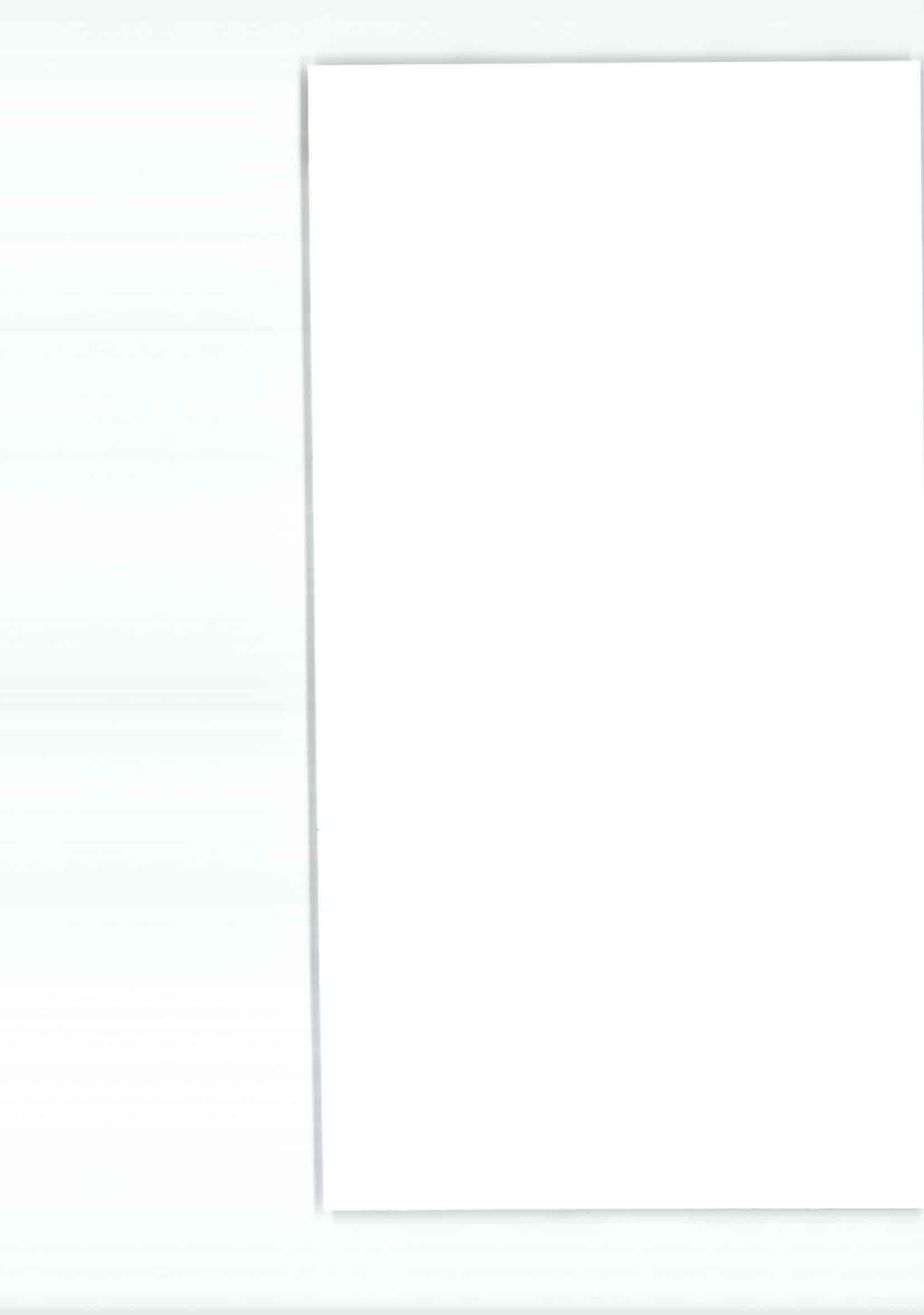
המקצוע

125

סמסמ

204 תאריך המבחן

	7	בגד
	8	נדסח
	9	



המחלקה 372 הנדסת מערכות מידע סמל המקצוע: קבוצה: 1

נוב

המקצוע: אחזור מידע וספריות דיגיטליות 1. 4406 0-372

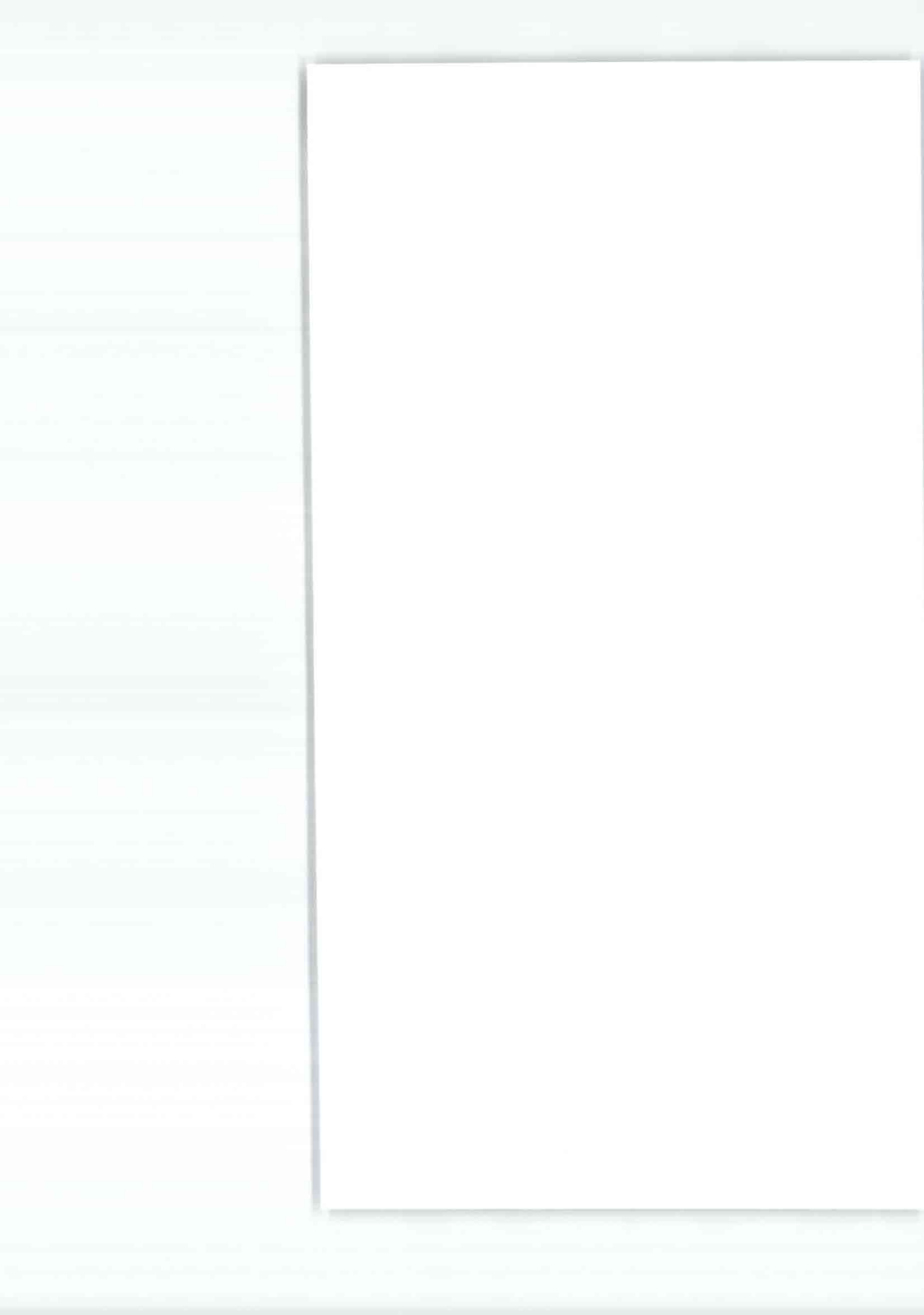
שם הסורה: ד"ר ב. שפירא 125

סמסטר: א-2011 מועד: א' אונברסיטת בן-גוריון

דע

תאריך המבחן: 31.01.2011 שעה: 13:30 דף: 1

מס' זהות	מס' נבחן	מס' סידורי	ציון סופי	ציון	ציון
21514484		55/1			חד מחלק
21551924	✓	56/2		80	חד מחלק
21565478		57/3			חד מחלק
21582333	✓	58/4		88	חד מחלק
21590831	✓	59/5		60	חד מחלק
21655477	✓	60/6		81	חד מחלק
21660394	✓	61/7		59	חד מחלק
21692215	✓	62/8		90	חד מחלק
21958715	✓	63/9		98	חד מחלק
26490748	✓	64/10		91	חד מחלק
26537522	✓	65/11		49	חד מחלק
32473258	✓	66/12		58	חד מחלק
32511230	✓	67/13		79	חד מחלק
32671406		68/14			חד מחלק
37110749	✓	69/15		65	חד מחלק
38059572	✓	70/16		83	חד מחלק
38080529		71/17			חד מחלק
39184783	✓	72/18		81	חד מחלק
46262721	✓	73/19		62	חד מחלק
60754629		74/20			חד מחלק
סה"כ					



כנגד

125

שם המורה: ד"ר ב. שפיידא

37

21 18-19-20

2011-N

18-19-20

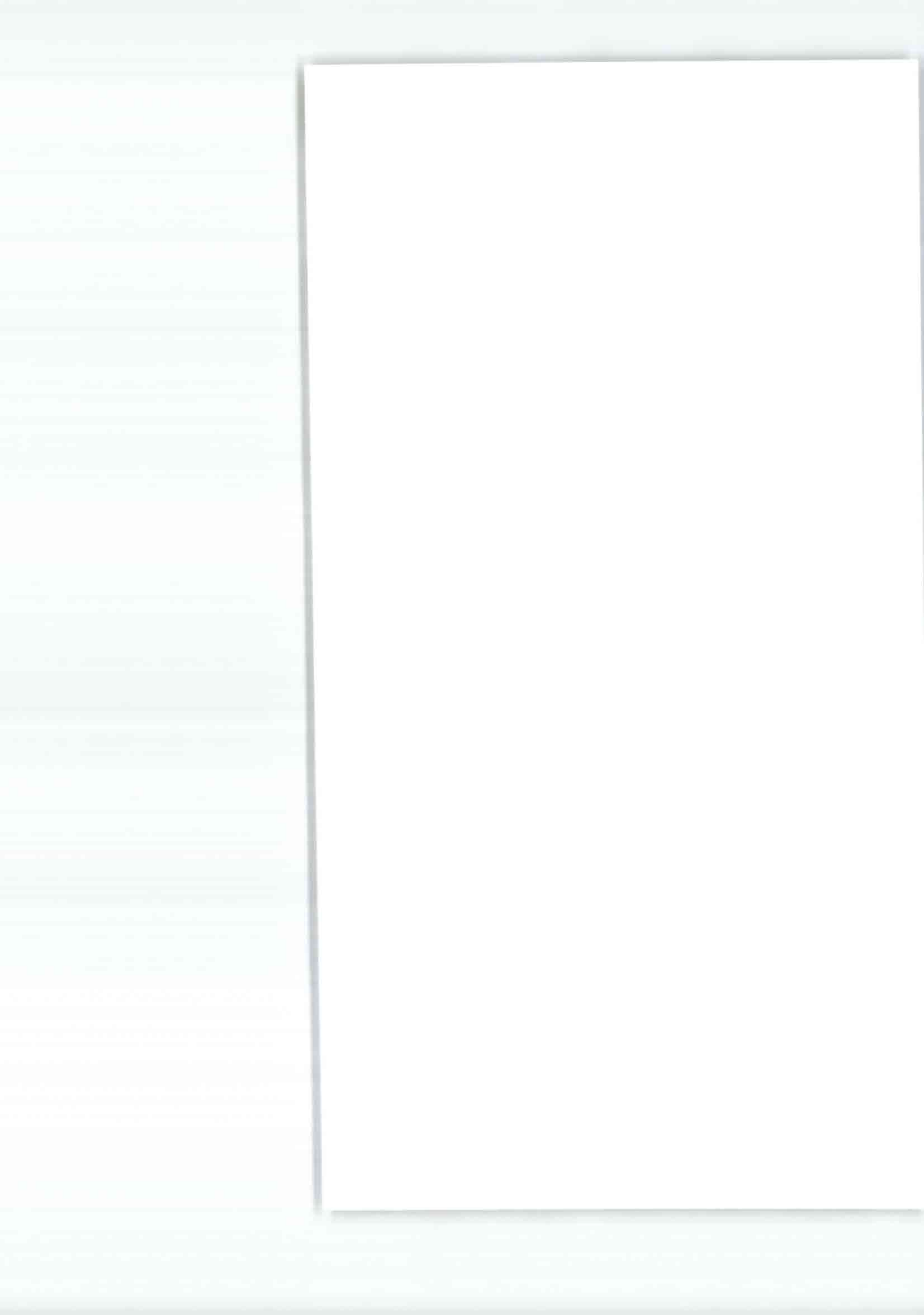
שעה : 13:3

31. 01. 2011

205-777,

27

[illegible]



נוב 372 הנדסת מערכות מידע המחלקה קבוצה 1: 7-12

המקצוע: מחזור מידע וספריות דיגיטליות 0-372.1.4406

שם המורה: ד"ר ב. שפירא 125

סמסטר: 2011-א 18-19-20 מועד: א' אונברסיטת בן-גור 21

תאריך המבחן: 31.01.2011 שעה: 13:30 דף: 1

מס' זהות	מס' נבחן	מס' סידורי	ציון סופי	ציון	ציון
32532491	✓	82/1		80	חד מחלקו
32568503	✓	83/2		89	חד מחלקו
34597625	✓	84/3		85	חד מחלקו
34653360	✓	85/4		76	חד מחלקו
36554012	✓	86/5		89	חד מחלקו
36591634	✓	87/6		88	חד מחלקו
36730158	✓	88/7		82	חד מחלקו
36822641	✓	89/8		72	חד מחלקו
36831394	✓	90/9		62	חד מחלקו
36850121	✓	91/10		79	חד מחלקו
36858595	✓	92/11		62	חד מחלקו
36863926	✓	93/12		83	חד מחלקו
36904357	✓	94/13		27	חד מחלקו
36939874		95/14			חד מחלקו
37061702	✓	96/15		80	חד מחלקו
37064144	✓	97/16		63	חד מחלקו
37775186	✓	98/17		56	חד מחלקו
37814993	✓	99/18		59	חד מחלקו
37963550	✓	100/19		74	חד מחלקו
37964731	✓	101/20		69	חד מחלקו

קבוצה: 1

סמל המקצוע:

372 הנדסת מערכות מידע

המחלקה:

אחזור מידע וספריות דיגיטליות 1. 4406 0-372

המקצוע:

ד"ר ב. שפירא

שם המורה:

125

מועד: מאי אוניברסיטת בן-גור

2011-א

סמסטר:

דע

18-19-20

דף 2

שעה: 13:3

31.01.2011

חדר-202 המבחן

ציון	ציון מבחן	ציון סופי	מס' סידורי	מס' מבחן	מס' זהות
חד מחלק	84		102/1	✓	37972189
חד מחלק	75		103/2	✓	37973831
חד מחלק	69		104/3	✓	37980471
חד מחלק	66		105/4	✓	38022067
חד מחלק	90		106/5	✓	38035226
חד מחלק	81		107/6	✓	38045084
חד מחלק	80		108/7	✓	38049482
חד מחלק	93		109/8	✓	38802211
חד מחלק	82		110/9	✓	39054580
חד מחלק	69		111/10	✓	39135017
חד מחלק	83		112/11	✓	39210109
חד מחלק			113/12		39399449
חד מחלק	79		114/13	✓	39532775
חד מחלק	70		115/14	✓	39540786
חד מחלק	62		116/15	✓	39730379
חד מחלק	71		117/16	✓	39780630
חד מחלק	66		118/17	✓	39877899

