

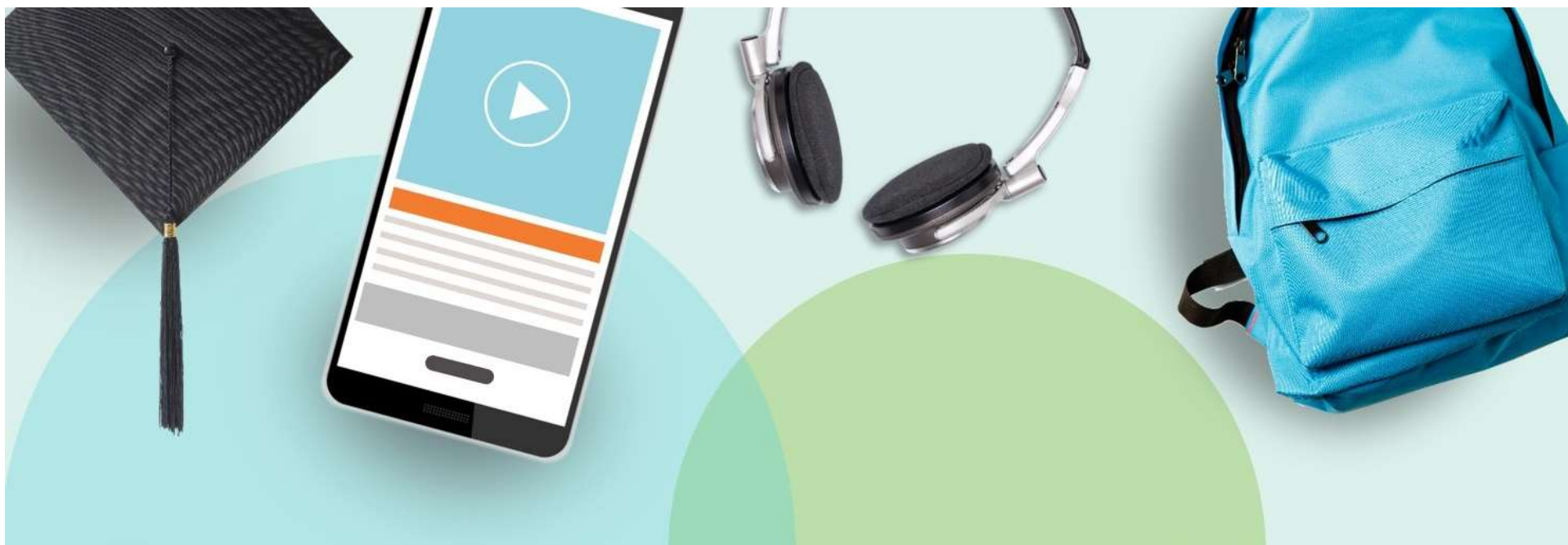
מנוע החיפוש בפרימו - מאחורי הקלעים של החיפוש בעברית

מרדכי גורן מפתח בכיר בפרימו



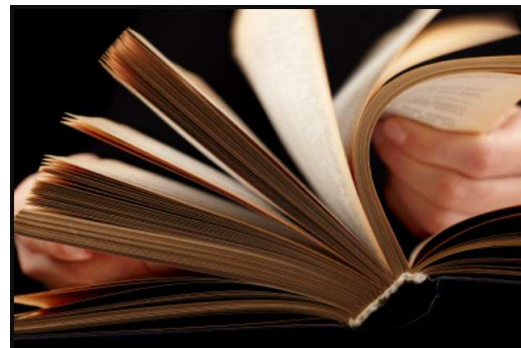
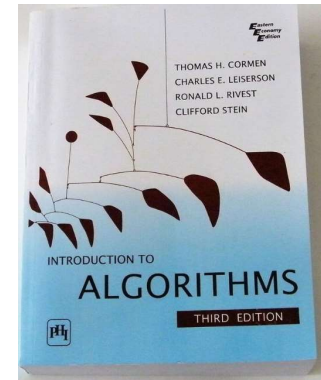
על סדר היום

- חיפוש כללי
- אתגרים בעברית ופתרונות
- דוגמאות מנותחות



תהליך אינדקס וחיפוש מקצה לקצה

חיפוש בספר



תוכן העניינים	
3	הקדמה
5	תוכן העניינים
6	מבוא
	התפתחות הכלים וכלליהם; האוטומטון; מחשבים הראשונים; הויכוח על רמת המחשב; הבעיה המועילית לדיון.
11	פרק א: מבנה תוכנית השחמט
	יצוג נתונים; איתחול התוכנית; גוף התוכנית; סיום התוכנית; שימושים אחרים בתוכנית.
17	פרק ב: החשיבה של המחשב
	רקע לשגרה 'בחר מהלך'; עיקרון מינו מקס; עיקרון נמקס; מבנה שגרת החשיבה של המחשב; איטרטיבה מעמיקה; בעיית היעילות; טבלת מצבים; גיוס אלפא בטא; תופעת האופק; יצירת מהלכים וזקרים; פונקציית הערכת מצב; סיכום

מסמך 123



ExLibris
a ProQuest Company

כותר: מנהגי הכשרות

מחבר: שרון שלום

נושא: יהודי אתיופיה

כשרות ; שחיטה

תיאור: עולם הכשרות ...

אינדקס (תוכן עניינים)

- אתיופיה ⏪ מסמך 123, מסמך 543, ...
- הכשרות ⏪ מסמך 11, מסמך 123, ...
- יהודי - ⏪ מסמך 123, מסמך 234, ...
- מנהגי - ⏪ מסמך 50, מסמך 123, ...



חיפוש הופך לשאילתה לוגית

חיפוש:

מנהגי הכשרות

שאילתה לוגית:

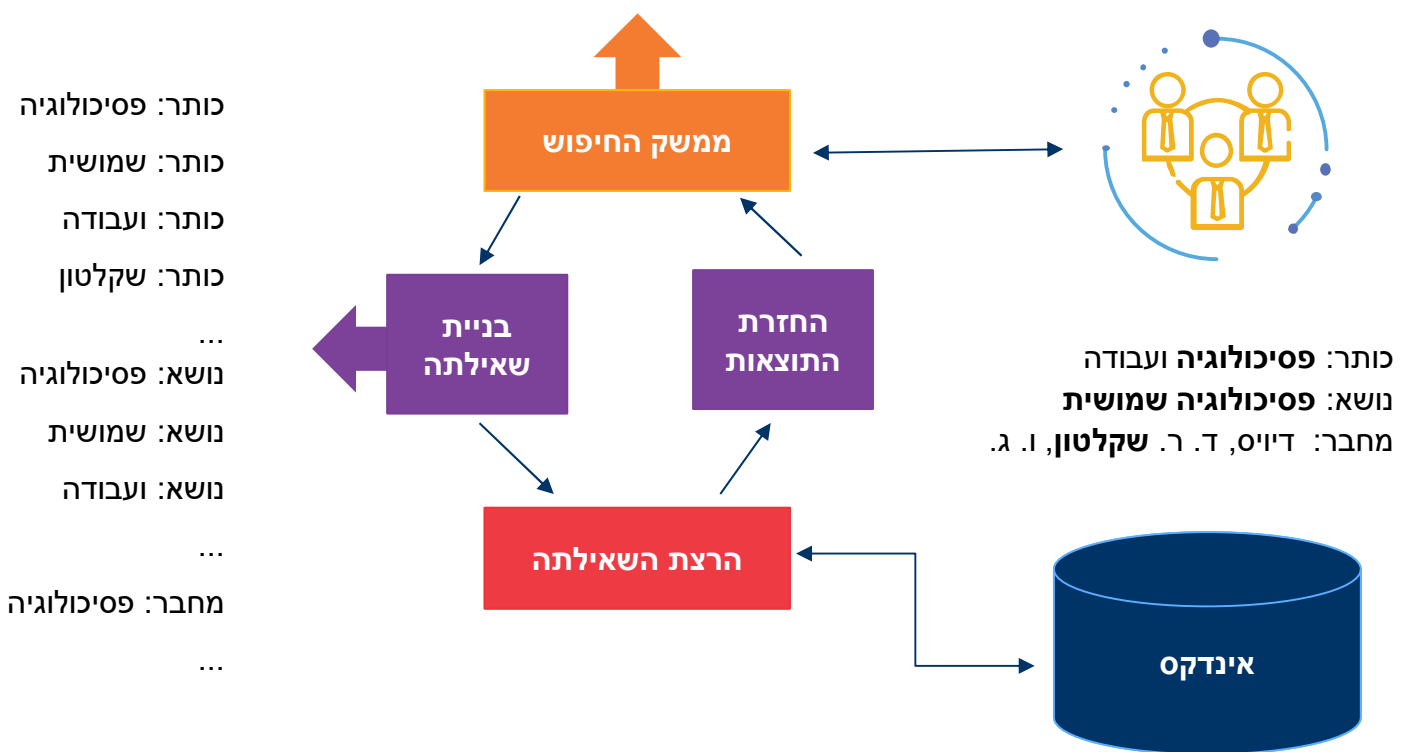
כותר: מנהגי או מחבר: מנהגי או ...

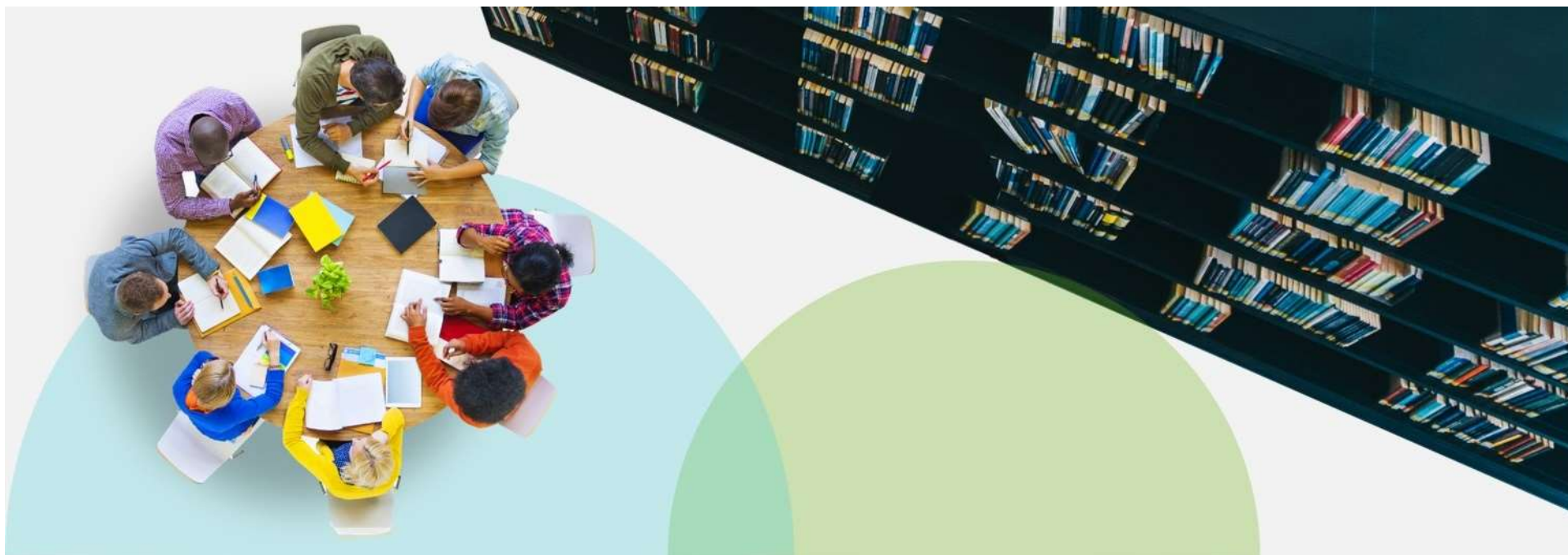
וגם

כותר: הכשרות או מחבר: הכשרות או ...

חיפוש

"פסיכולוגיה שמושית ועבודה, שקלטון"





אלגוריתם החיפוש

שלושה דגשים





- כל מילה באיזשהו שדה
- מספר התוצאות

מנהגי הכשרות

עמוד 1 95 תוצאות בהתאמה אישית



שאלתה לוגית:

כותר: מנהגי או מחבר: מנהגי או ...

וגם

כותר: הכשרות או מחבר: הכשרות או ...

מסמך 789

כותר: מנהגי חנוכה

מחבר: נילי בן עזרא

נושא: חנוכה; מנהגים

תיאור: ... כשרות הסופגניות...

מסמך 987

כותר: פורים וקפצונים

מחבר: מרדכי גורן

נושא: מנהגים; פורים; משלוחי מנות; כשרות

תיאור: מנהג ... פורים ... סכנות ... כשרות





- באיזה סדר יופיעו התוצאות
- ביטוי לפני מילים בודדות
- כותר לפני נושא לפני תוכן עניינים



4 ספר
נתיבות הכשרות : דינים ומנהגים לבני אשכנז וספרד
וינברגר, יעקב; זיכל, מאיר; בר שאול, מלכיאל; תשנ"א
נמצא בספרייה - הקישו על מיקום <

5 ספר
נתיבות הכשרות : הכשרות בבית. דינים ומנהגים לבני אשכנז וספרד.
בר שאול, מלכיאל; וינברגר, יעקב; זיכל, מאיר
נמצא בספרייה - הקישו על מיקום <

6 מאמר
חגבים - ממאכלי השולחן הכשר של יהודי תימן.
אבישר, ינון; 1994
תהודת כשרות. 1. 52-53, ניסן תשנ"ד.
לבדיקת הכרכים/עותקים פנה לספרייה <

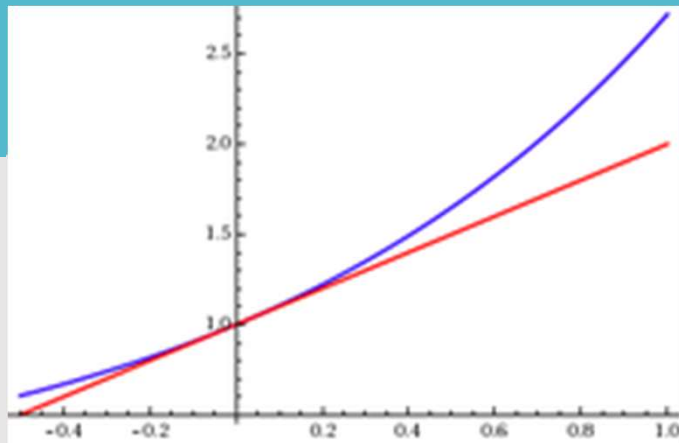
7 ספר
ספר כשרות בפסח: / כולל הלכות ומנהגי פסח בענייני כשרות המזון לפסח
אדרעי, עמרם; תשס"ד
נמצא בספרייה - הקישו על מיקום <

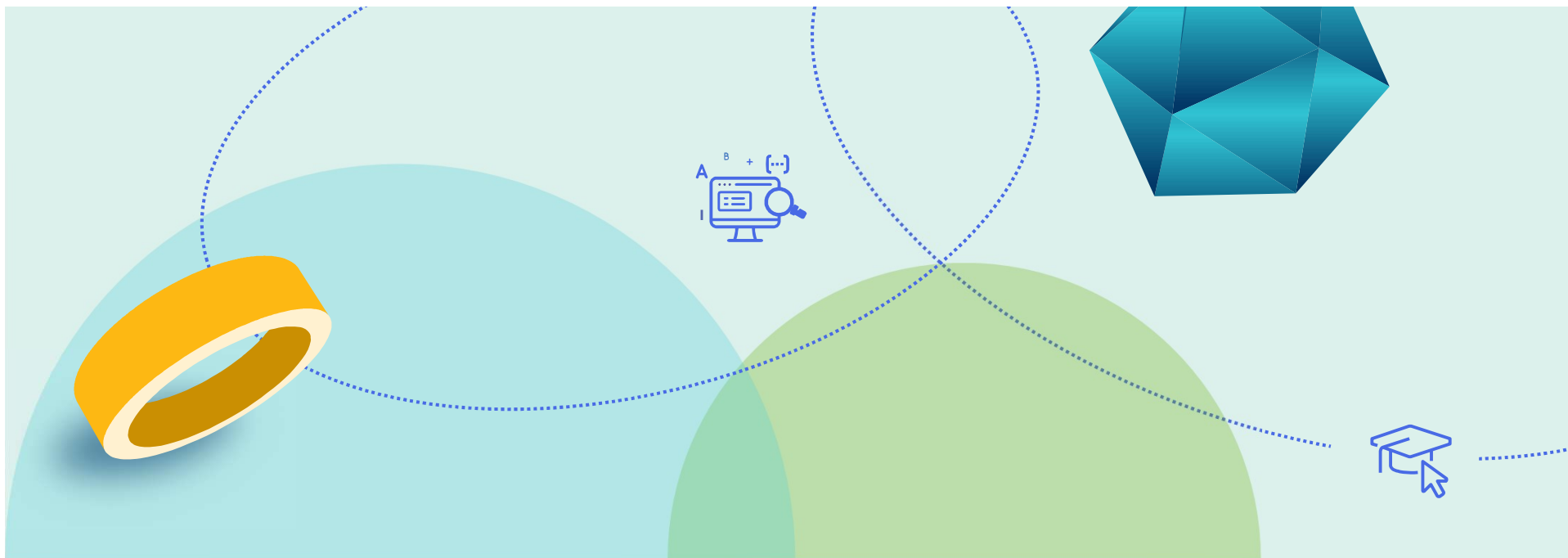


קירוב – בין איחזור לדירוג



- מה אם יש טעות כתיב?
- מה אם יש ארבע מתוך חמש מילים?
- קיימים במערכת קירובים
- משפיע על דירוג ואיחזור

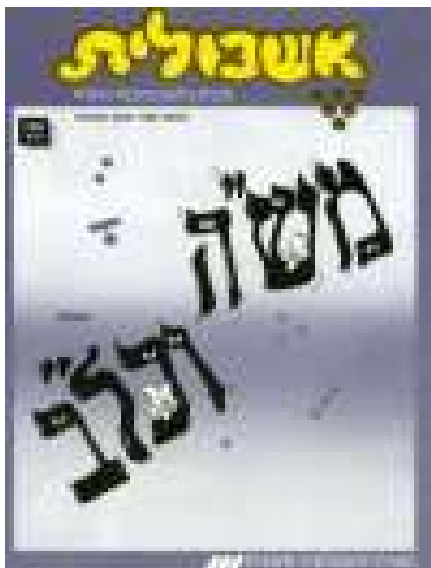




האתגר בעברית

האתגר בעברית

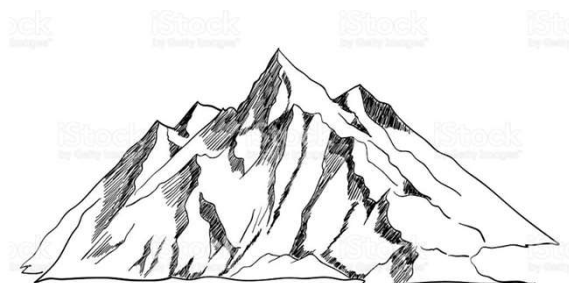
- הטיות
- מלא חסר
- דו משמעות



הטיות



- וּלְדָן אָמַר דָּן גּוֹר אֶרֶיָּה יִזְנֶק מִן הַבְּשָׁן [דברים לג 22]
- באנגלית: **And To Dan**
- יש גם מחבר וּלְדָן



- מהר: מהירות ? מן ההר?
- באנגלית
- From The Mountain
- Fast

מלא/חסר

- מילדות: מיילדות (מלא) ? מן הילדות?
- באנגלית:
 - Since Childhood
 - Midwives

ראשי תיבות – דומה למלא חסר

- רמטכ"ל צה"ל
- רמטכל צהל



דו משמעות

- מהי דלת?
- זוהי דלת שפותחת את הפל

- ולהבדיל, מלשון דל
לא נשאר זולת דלת עם הארץ
(מלכים ב כד14)





הפתרון בעברית

גלגולו של פתרון



Synonym

HebMorph

ראשי תיבות

מלינגו



שנות האלפיים

2014

2017





ראשי תיבות

ראשי תיבות

- רמטכ"ל צה"ל
- רמטכל צהל



אינדקס (תוכן עניינים)



• צה"ל < מסמך 50, מסמך 123, ...

• רמטכ"ל < מסמך 123, מסמך 543, ...

• צהל < מסמך 234, מסמך 456, ...

• רמטכל < מסמך 11, מסמך 456, ...



צד חיפוש – מילה אחת



- חיפוש: רמטכ"ל
- רמטכ"ל או רמטכל
- עדיפות לצורה רמטכ"ל

- חיפוש: רמטכל
- זיהוי רמטכל צריך גרשיים – ע"פ רשימה סגורה

- רמטכל או רמטכ"ל
- עדיפות לצורה רמטכל

צד חיפוש - יותר ממילה אחת



- חיפוש: רמטכ"ל צה"ל
- רמטכ"ל צה"ל או רמטכל צהל
- עדיפות לצורה רמטכ"ל צה"ל

- חיפוש: רמטכל צהל
- רמטכל צהל או רמטכ"ל צה"ל
- עדיפות לצורה רמטכל צהל



מלינגו

פרימו משתמש במלינגו בפונקציונליות הבאה

2

מציאת צורות
שורש אפשריות

1

הכרעה באיזה
שורש לבחור

ייצוג בעזרת מזהה Hash

הדלת, דלת, דלתות < 766zzhe

מופעל בשדות: כותר, מחבר, נושא, תיאור

מופעל גם עבור ה-Resource Recommender



אינדקס (תוכן עניינים)



• דלת ⏪ מסמך 50, מסמך 123, ...
• ...

• דלתות ⏪ מסמך 14, מסמך 52, ...
• ...

• הדלת ⏪ מסמך 234, מסמך 456, ..., ...

• 766zzhe ⏪ מסמך 14, מסמך 52, מסמך 123, מסמך 234, מסמך 456, ...

חיפוש מילה אחת

חיפוש: דלת

766zzhe: כותר: דלת או כותר: 766zzhe

או

766zzhe: נושא: דלת או נושא: 766zzhe

או

חיפוש: הדלת

766zzhe: כותר: הדלת או כותר: 766zzhe

או

766zzhe: נושא: הדלת או נושא: 766zzhe

או

דלת: TOC

≠

הדלת: TOC

חיפוש יותר ממילה אחת

מצב קיצון

עד מרחק 8

22958zzhe: כותר: מצב או כותר:

52289zzhe: כותר: קיצון או כותר: וגם

...

עד מרחק 8

22958zzhe: נושא: מצב או נושא:

52289zzhe: נושא: קיצון או נושא: וגם

TOC: קיצון

וגם

TOC: מצב



כותר: פירוש המשניות של ח"נ ביאליק
מאמר ביקורת

מחבר: לחובר פישל

נושא: < ביאליק חיים נחמן 1873-1934
< משנה זרעים סדר
< פרשנות

מסמך 789

כותר: מנהגי חנוכה

מחבר: < נילי בן עזרא < מרדכי גורן

נושא: חנוכה; מנהגים

תיאור: ... כשרות הסופגניות...

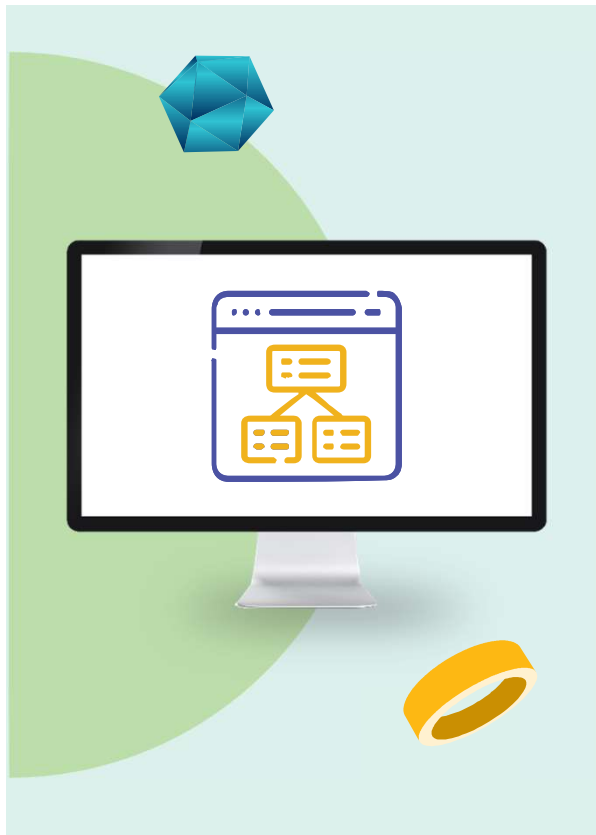
מסמך 987

כותר: פורים וקפצונים

מחבר: < מרדכי גורן < נילי בן עזרא

נושא: מנהגים; פורים; משלוחי מנות; כשרות

תיאור: מנהג ... פורים ... סכנות ... כשרות



חיפוש יותר ממילה אחת

מצב קיצון

עד מרחק 8

22958zzhe: כותר: מצב או כותר:

52289zzhe: כותר: קיצון או כותר: וגם

...

עד מרחק 8

22958zzhe: נושא: מצב או נושא:

52289zzhe: נושא: קיצון או נושא: וגם

TOC: קיצון

וגם

TOC: מצב



כותר: פירוש המשניות של ח"נ ביאליק
מאמר ביקורת

מחבר: לחובר פישל

נושא: (0) ביאליק חיים נחמן 1873-1934

אחת שתיים שלוש

(8) משנה זרעים סדר

אחת שתיים שלוש

(14) פרשנות

האם התכוונת לחפש את: ספרות ישראל?

0 נבחרו / עמוד 1 / 581 תוצאות

הצורך בספריות אזוריות בישראל : סקר
 Z843.I8 T9
 יד אדמונד ג'ימס דה רוטשילד
 ירושלים : יד אדמונד ג'ימס דה רוטשילד תשכ"ד
 נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם



0 נבחרו / עמוד 1 / 1,862 תוצאות

מעמדה של מדינת ישראל והשפעתה על יהדות אנגליה / ישראל פיינשטיין
 DS135.E5 F455
 פיינשטיין, ישראל
 ירושלים : האוניברסיטה העברית, המכון ליהדות זמננו תשמ"ד 1983
 נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם



חיפוש יותר ממילה אחת עם מחבר

מצב קיצון כהן



- קיים מנגנון לזיהוי מחברים
- מחבר: **כהן** וגם **מצב קיצון**
- מפעילים הרחבות מיוחדות של מלינגו על 'כהן'

מבוא לכימיה

אוסף הספרייה / ספרים ועוד / X

הכול / עם מילות החיפוש / בכל מקום

0 נבחרו / 3 תוצאות

1

מבוא למבנה החומר : פרקי פיסיקה וכימיה : לתלמיד
 LTQC171 .I8
 ישראל. משרד החינוך והתרבות. היחידה לתוכניות לימודים
 ירושלים : חמו"ל תשכ"ט 1969
נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם

2

מבוא למבנה החומר : פרקי פיסיקה וכימיה : למורה
 LTQC171 .I8 1969
 ישראל. משרד החינוך והתרבות. היחידה לתוכניות לימודים
 ירושלים : חמו"ל תשכ"ט 1969
נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם

פעם התוצאות לא חזרו

חיפוש יותר ממילה אחת עם מחבר

מצב קיצון כהן

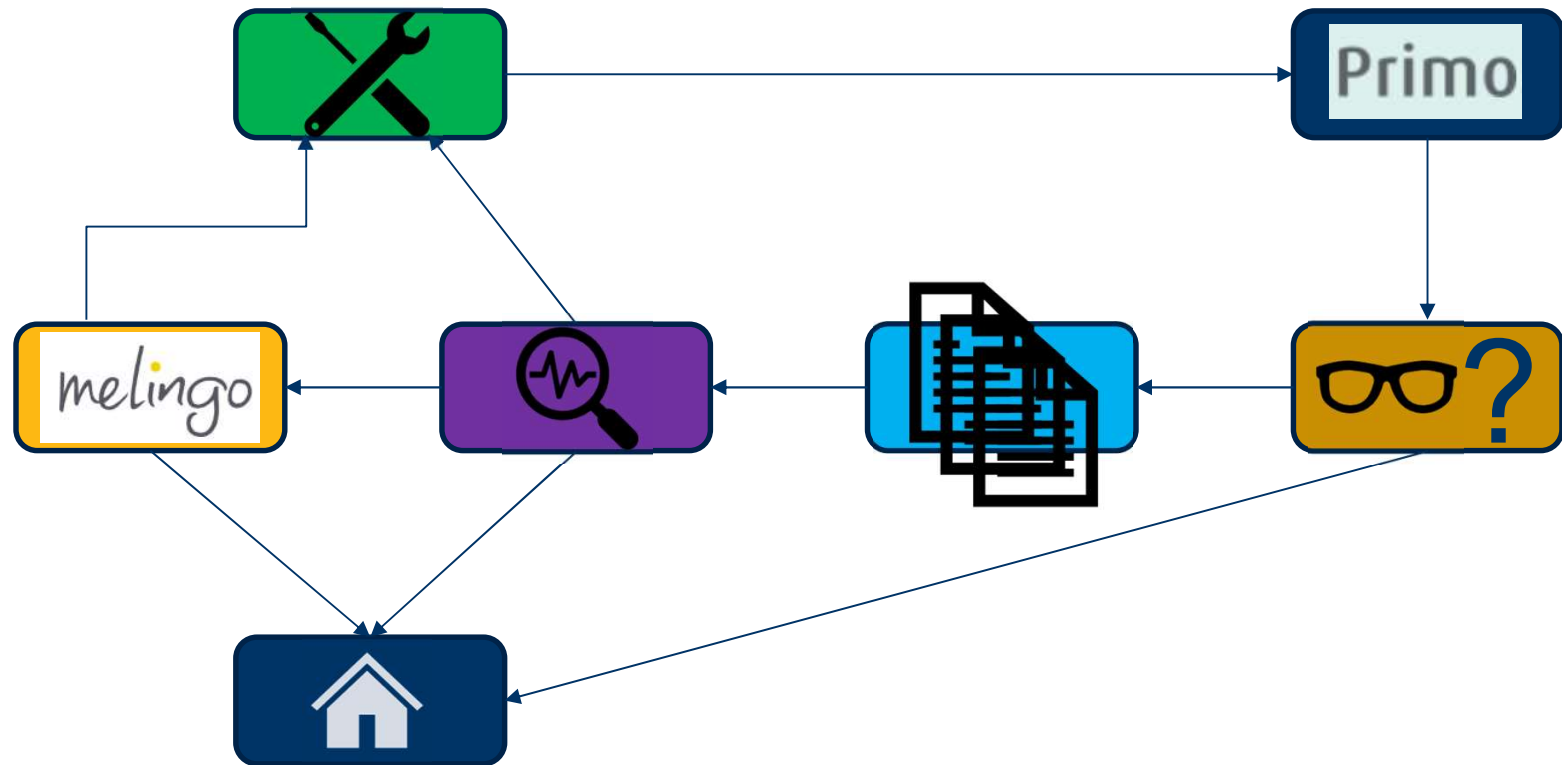


- קיים מנגנון לזיהוי מחברים
- מחבר: **כהן** וגם **מצב קיצון**
- מפעילים הרחבות מיוחדות של מלינגו על 'כהן'



- 'לכימיה' זוהה בתור מחבר
כי מאד נפוץ בשדה המחבר...
- עבור מחברים מלינגו מופעל אחרת
- תוקן בפרימו (אוגוסט 2019)

יש לכם שאלה? קריאה עד תיקון



From: Nili Natan

Sent: Monday, April 1, 2019 1:33 PM

To: MELI - Primo Forum in Israel <IL-MELI-PRIMO@LISTSERV.TAU.AC.IL>

Subject: שיפורים בחיפוש בעברית בפרומו - רכיב מלינגו שודרג בהצלחה



Primo News

ExLibris
a ProQuest Company

משתמשי פרימו יקרים,

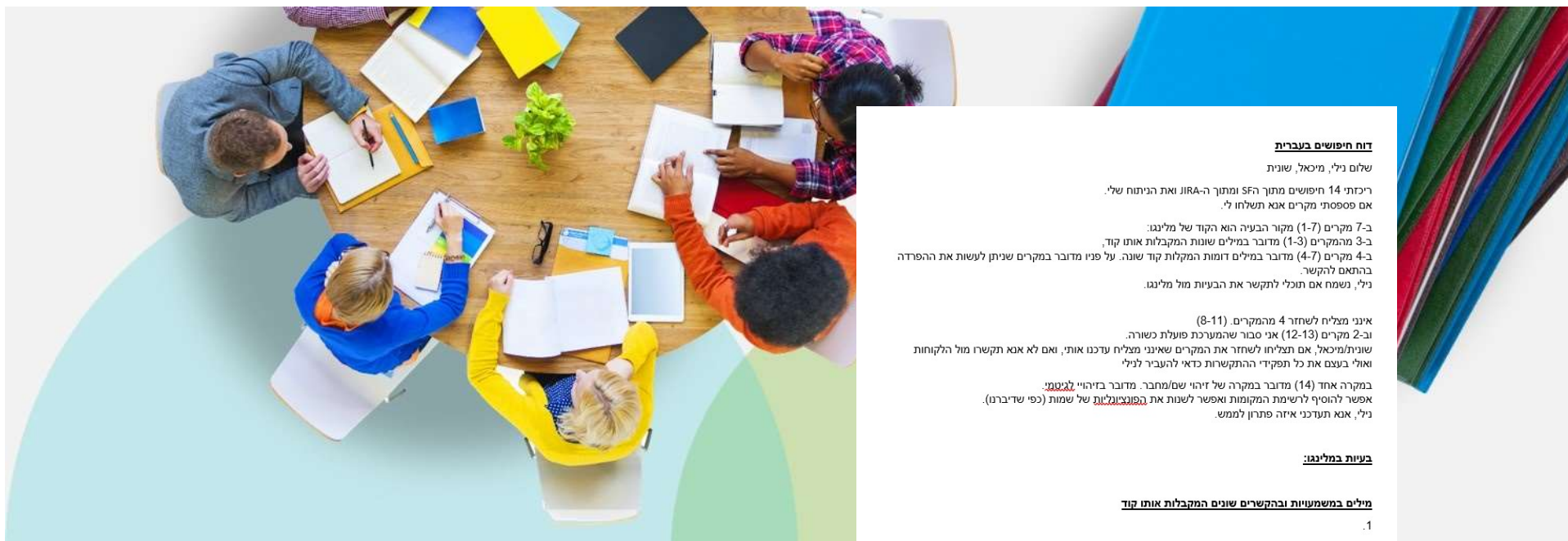
אנחנו שמחים לבשר שתהליך האינדוקס המלא הסתיים ורכיב מלינגו בפרומו שודרג בהצלחה. זו הזדמנות נוספת להודות לנציגי קבוצת העבודה של פרימו על הדחיה ושיתוף הפעולה.

להלן השיפורים בחיפוש בעברית בעקבות שדרוג רכיב מלינגו:

- **תיקונים למילון של מלינגו**
לדוגמה תיקון בזהיו המילה בלגרד כעיר בירה (בעבר זוהה כחלק מהשורש גרד, גרדה, מגרד, לגרד)
- **זיהוי צירופי מילים:** שיפור הניתוח המורפולוגי של מלינגו על ידי איתור זיהוי צירופי מילים במהלך האינדוקס והחיפוש בפרומו, לדוגמה
 - שובר גלים – נותח קודם כשובר (קופון)
 - נשאי תערובת – מנותח כנישואין - נותח קודם נשאי (תחביר)
 - פתוח עירוני מנותח כפיתוח עירוני – לפני השיפור נותח כפתוח
- **כתיב מלא חסר:** מתן דירוג גבוה יותר לכתיב מלא חסר. לדוגמה שעמום – מנותח כשיעמום - לפני השיפור הוא נותח +שעמום

בכל שאלה או הבהרה או משוכ ניתן לפנות קריאה באמצעות מערכת ה- Salesforce לצוות התמיכה

בברכה
אקסליבריס



דוח חיפושים בעברית

שלום נילי, מיכאל, שונית

ריכזתי 14 חיפושים מתוך SFN ומתוך ה-JIRA ואת הניתוח שלי. אם ספספסתי מקרים אגא תשלחו לי.

ב-7 מקרים (1-7) מקור הבעיה הוא הקוד של מלינגו:

ב-3 מהמקרים (1-3) מדובר במילים שונות המקבלות אותו קוד,

ב-4 מקרים (4-7) מדובר במילים דומות המקלות קוד שונה. על פניו מדובר במקרים שניתן לעשות את ההפרדה בהתאם להקשר.

נילי, נשמח אם תוכלי לתקשר את הבעיות מול מלינגו.

איני מצליח לשחזר 4 מהמקרים (8-11)

וב-2 מקרים (12-13) אני סבור שהמערכת פועלת כשורה.

שונית/מיכאל, אם תצליחו לשחזר את המקרים שאיני מצליח עדכנו אותי, ואם לא אגא תקשרו מול הלקוחות ואולי בעצם את כל תפקידי ההתקשרות כדאי להעביר לנילי

במקרה אחד (14) מדובר במקרה של זיהוי שם/מחבר. מדובר בזהוי לגיטמי:

אפשר להוסיף לרשימת המקומות ואפשר לשנות את הפונקציות של שמות (כפי שדיברנו).

נילי, אגא תעדכני איה פתרון לממש.

בעיות במלינגו:

מילים במשמעויות ובהקשרים שונים המקבלות אותו קוד

.1

דוגמאות בהם הגורם הוא פרימו



כותר: שביל ישראל - שבילים ארוכים בישראל

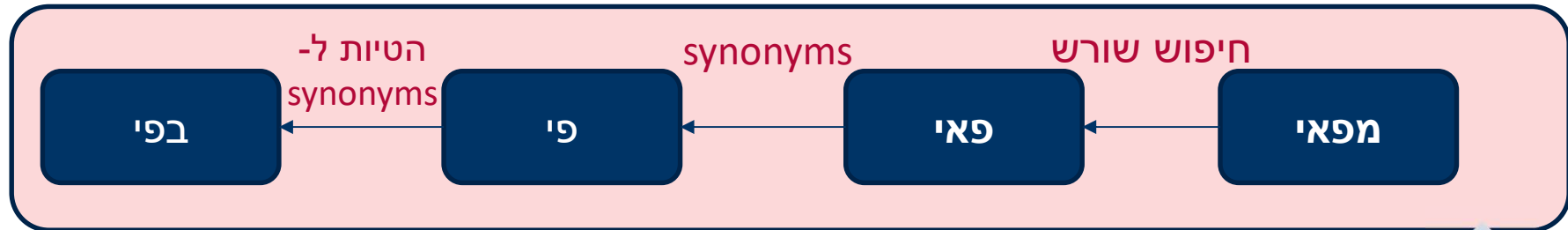
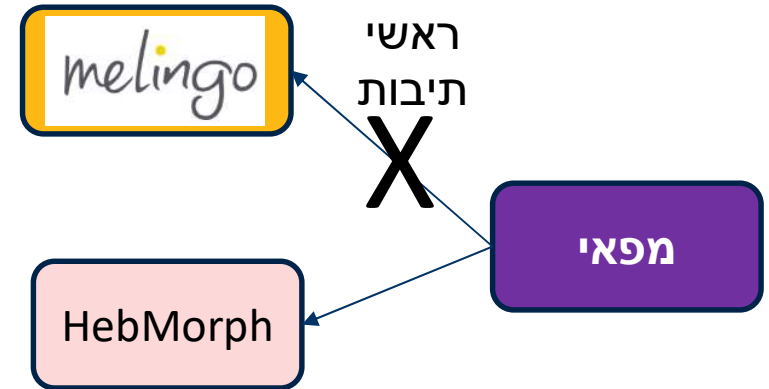
מחבר: סער יעקב

תיאור: מיום שראה אור בפי ... המפות
במדריך ... רכב שטח על מנת להגיע לתחילתו
או סופו של כל קטע.

איך מפאי הופך ל-בפי (לא נעים...)

תוקן בפרימו (פברואר 2020)

- תיקון מנגנון מציאת שורש
- הקשחת מנגנון הפעלת הטיות על סינונים

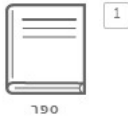




דוגמאות בהם הגורם הוא מלינגו

...   

ספר מהרי"ל : שחיבר ותיקן מנהגי ק"ק אשכנז... ונתוסף עליו חדושים רבים
BM700 .Y3 1959



אמנם משמעות שונה, אבל המערכת בסדר

יעקב בן משה, הלוי, 1427-?1365
בני-ברק : המאור תשי"ט 1959

נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם

...   

פרוש ר' אברהם אבן עזרא על תרי-עשר : כתב יד ותיקן 75 / הקדמה ומפתח מאת
איתן לוין



מצוין

BS1360 .I23 1976

אבן עזרא, אברהם בן מאיר, 1089-1164; לוין, איתן
ירושלים : מקור תשל"ו 1976

נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם

...   

The Vatican : 64 magnificent Photopages of the sights and scenes of Vatican
City and its treasures / [edited by Edward M. Kinney from photos ; taken by
Francesco Giordani]



יפה. (בזכות שדות נוספים)

DG800 .G45

Giordani, Francesco, 1896-; Kinney, Edward M; Vatican Pavillon (New York World's Fair) Inc. U.S.A;
New York : Vatican Pavillon, New York World's Fair c1964

נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם

מודל הוותיקן ושאלת ירושלים / אמנון רמון, רותם גלעדי ; בהנחיית ראובן מרחב
 DS109 .R33 1997
 רמון, אמנון; גלעדי, רותם; מרחב, ראובן; מכון ירושלים למחקר;
 ירושלים : מכון ירושלים לחקר ישראל 1997

חוזר במקום ה-100.
לא חוזר אם מחפשים בכותר בלבד



העדפה **מתקנת** בישראל : הגדרת מדיניות והמלצות לחקיקה / הילי מודריק-אבן חן ;
 בהנחיית מרדכי קרמניצר ודוד נחמיאס
 KA766.A1 M7 2000
 מודריק-אבן חן, הילי, 1971-; המכון הישראלי לדמוקרטיה
 ירושלים : המכון הישראלי לדמוקרטיה 2000

חוזר במקום ה-10.
לא צריך לחזור כלל



תגובת מלינגו


ותיקן, וותיקן - מקבלים קוד שונה zzhe11447 לעומת 40461zzhe
 בחיפוש של "ותיקן" נמצא הפועל "תיקן" בסביחות גבוהה יותר.


אנחנו מניחים שזה קרה בשאלתא, אם אכן כך הדבר, אנחנו נתקן את זה כך ש"ותיקן" (המקום) יקבל עדיפות (גם באופן פרטי במקרה הזה, וגם נבדוק אם יש בעיה במקרים דומים, היות שאכן פועל עם וו החיבור לא אמור להילקח כאפשרות הראשונה בחיפוש כאשר יש אפשרויות אחרות).

מפוטרים

אוסף הספרייה / ספרים ועוד / X

הכול עם מילות החיפוש בכל מקום

1  **פטר קרופוטקין : תולדותיו, רעיונותיו וספריו : שלושים שנה למותו (1921-1951) / מאת א' גורל**
 HX914.K7 G6 1951
 גורל, א
 תל אביב : חמו"ל תשי"א 1951
 נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם

2  **הארי פוטר והילד המקולל : חלקים ראשון ושני / מחזה חדש מאת ג'ק ת'ורן ; מאנגלית: גילי בר-הלל סמו ; ערך: אמיר צוקרמן**
 PR6120.H67 H3716 2016
 ת'ורן, ג'ק; רולינג, ג'יי. קיי., 1965-; טיפאני, ג'ון, 1971-;
 תל אביב : ספרי עליית הגג 2016
 מושאל

תגובת מלינגו

מפוטרים, פוטר (בהקשר של הארי פוטר), פטרה
 מקבלות אותו קוד:
 zzhe11561

השם "הארי פוטר" מופיע אצלנו במאגר של שמות פרטיים, אך לא במאגר המילוני. אנחנו בתהליך הוספת מאגר השמות הפרטיים למודול החיפוש, שישפר מקרים כמו זה.



דוגמאות שהמערכת בסדר

איראן ומערכת היחסים הבין-ערביים / ג. בן-דור
DS274 .B4

בן-דור, גבריאל, 1944-; המכון ללמודי המזרח התיכון
חיפה : מלאת 1976

נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם



190

4

בין ערביים / משה פומרק
DS125.3.P6 A3

פומרק, משה
תל-אביב : אותפז תשל"א 1971

נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם



190

1

מערכת היחסים הבין ערביים והסכסוך הערבי-ישראלי / ג. בן-דור
DS119.7 .B3657

בן-דור, גבריאל, 1944-; המכון ללמודי המזרח התיכון
חיפה : מלאת 1976

נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם



190

5

הרמוניה בין ערביים / דוד (דודו) אליעז
PJ5055.2.L475 H37 2007

אליעז, דוד
ח"מ : המחבר תשס"ז 2007

נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם



190

2

מי בעד שוויון? : שוויון בין ערבים ליהודים בישראל : סיכום מחקר עמדות
נוהאד עלי ושי ענבר ; עריכה: רון גרליץ ויסמין הלוי
DS113.7 .A5256 2011

עלי, נוהאד ענבר, ושי, ענבר, עריכה: רון גרליץ ויסמין הלוי. "המכון ללמודי המזרח התיכון" חיפה : מלאת 2011



190

אלקטרוני

6

עיר בין ערביים : לאומיות מזדקנת ביפו / חיים חזן, דניאל מ
DS110.Y3 H39 2011

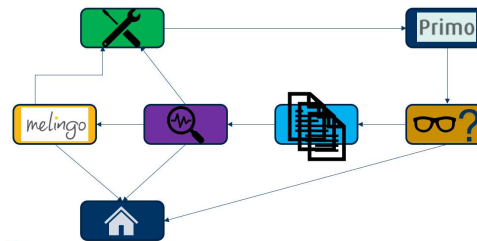
חזן, חיים; מונטרסקו, דניאל; מכון ון ליר (ירושלים);
תל-אביב : הקיבוץ המאוחד תשע"א 2011

נמצא ב אוסף כללי (14 ימים) ; בדקו זמינות עותקים נוספים ומיקומם



190

3



מה הלאה?

- ממשיכים
- נקיון קובץ הסינונים Q3 2020 (חיפה)
- VE – הוספת קובץ ראשי תיבות (acronyms) ✓
- שדרוג מלינגו
- בחינת הטמעה ב-CDI

סיכום

- חיפוש ואינדקס
- שיקולים מנחים
- סינונים וראשי תיבות
- מלינגו



סיפור – מהמורה בכיתה ו'

אבאלאבאבאבאב אלאבבבאבאבא



ב' אב תשפ"ו / 23/7/2020

אֲבָא לֹא אָבִיא בְּא' בְּאָב
אֶלֶּא בְּב' בְּאָב אָבִיא

תודה רבה!

