כלי לחיפוש בלשני בעברית

מהו חיפוש בלשני?

- CPQ is flexible and efficient query processor developed by IMS Open Corpus Workbench (CWB).
- It allows to perform linguistically motivated queries, e.g., search for a particular lemma followed by a word (verb?) in past tense.

- כפיר וחיים שללשוילי בהדרכת פרופ' שולי וינטנר מאוניברסיטת חיפה התאימו את CPQ לעברית
- דליה בוז'ן ואלון איתי בנו web interface

שמנו דגש הוא על קלות השימוש תוך ויתור על האופציה לחיפוש בוליאני .

כל שאילתה מורכבת מסדרה של עד 4 מילים. כל מילה ניתן לחפש על פי תמנית פני השטח (surface token), הערך הלקסיקלי (הלמה, הערך שמופיע במילון) ולגבי פעלים ניתן לחפש גם על-סמך השורש.

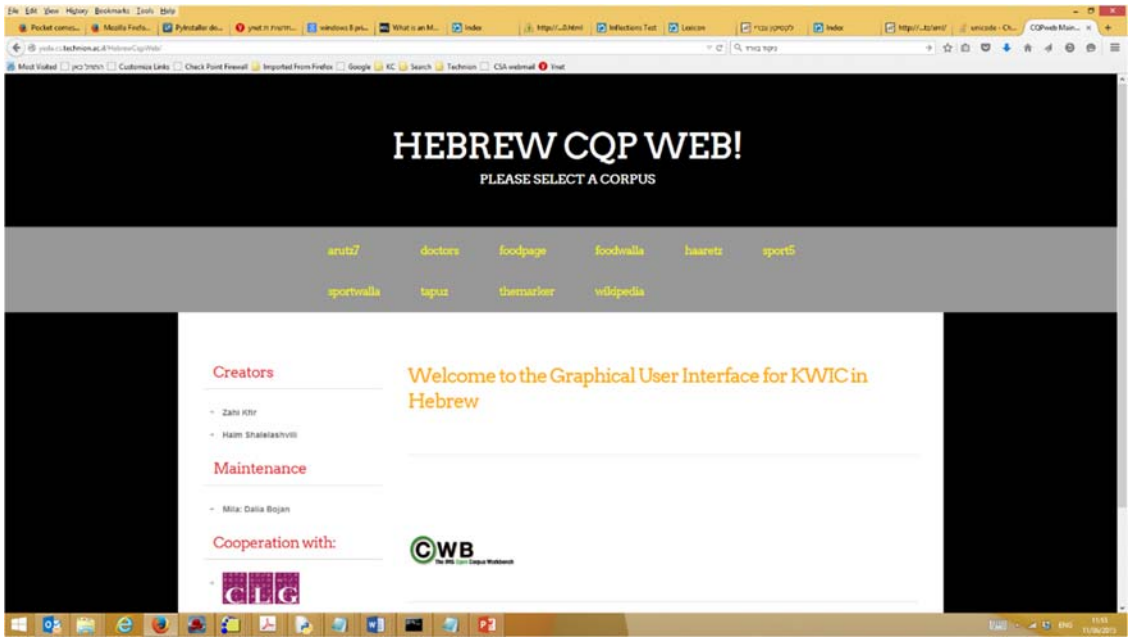
את התכנה ניתן להפעיל בקשורית הבאהhttp://yeda.cs.technion.ac.il/HebrewCqpWeb

בעיות נא לשלוח ל-mila@cs.technion.ac.il

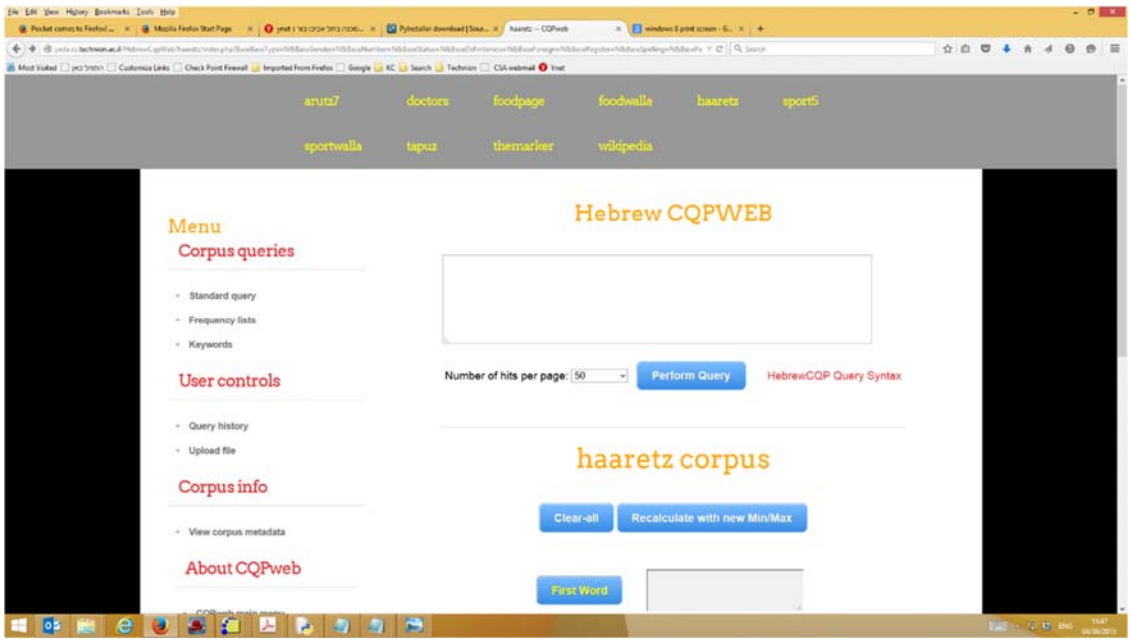
להלן הדגמה:

כניסה לאתר החיפוש דרך אתר מילה

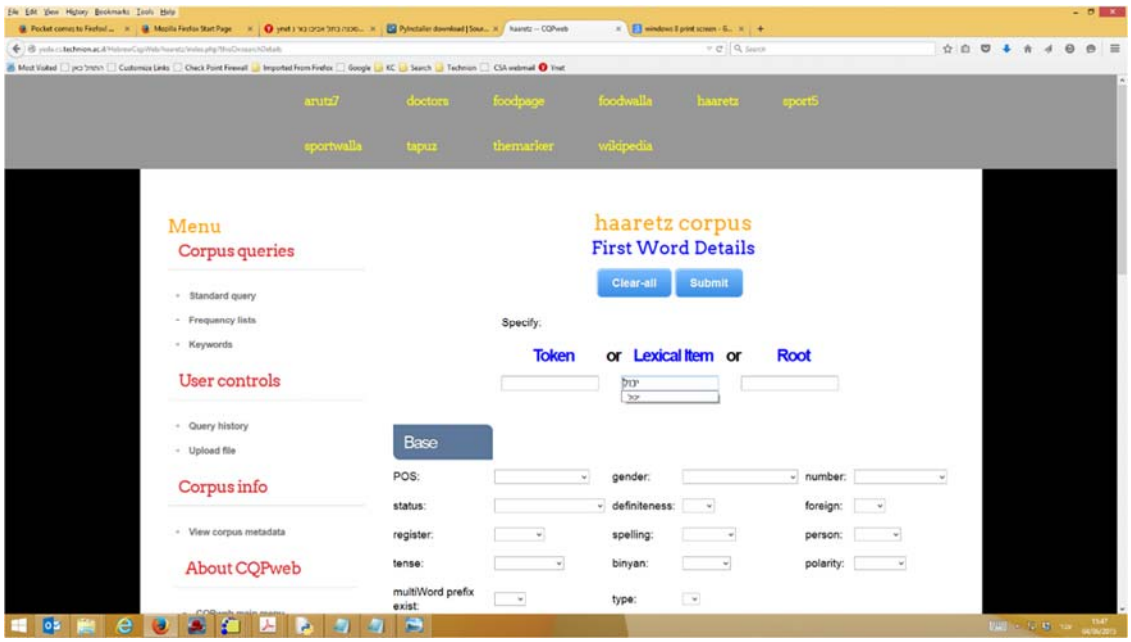
דף הפתיחה לחיפוש



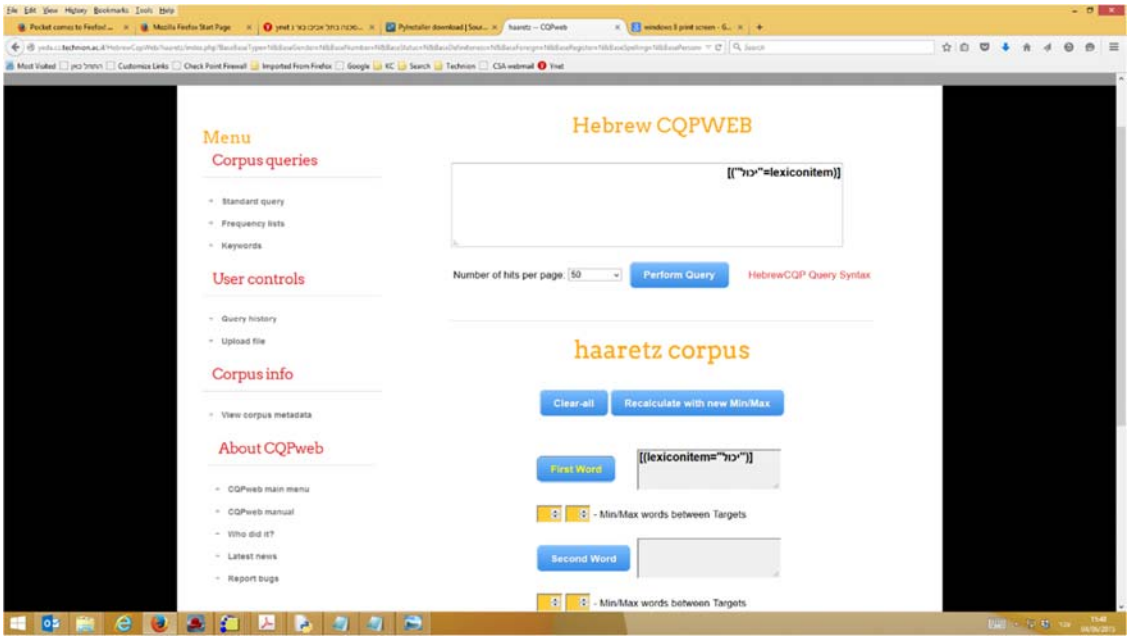
יש כמה קורפוסים. בדוגמא זו נבחר על אתר "הארץ". הקלקה על haaretz (למעלה בצהוב) תוליך אותנו לדף הבא:



נבחר את המילה הראשונה בשאילתה. הקלקת First word תוביל למסך הבא:

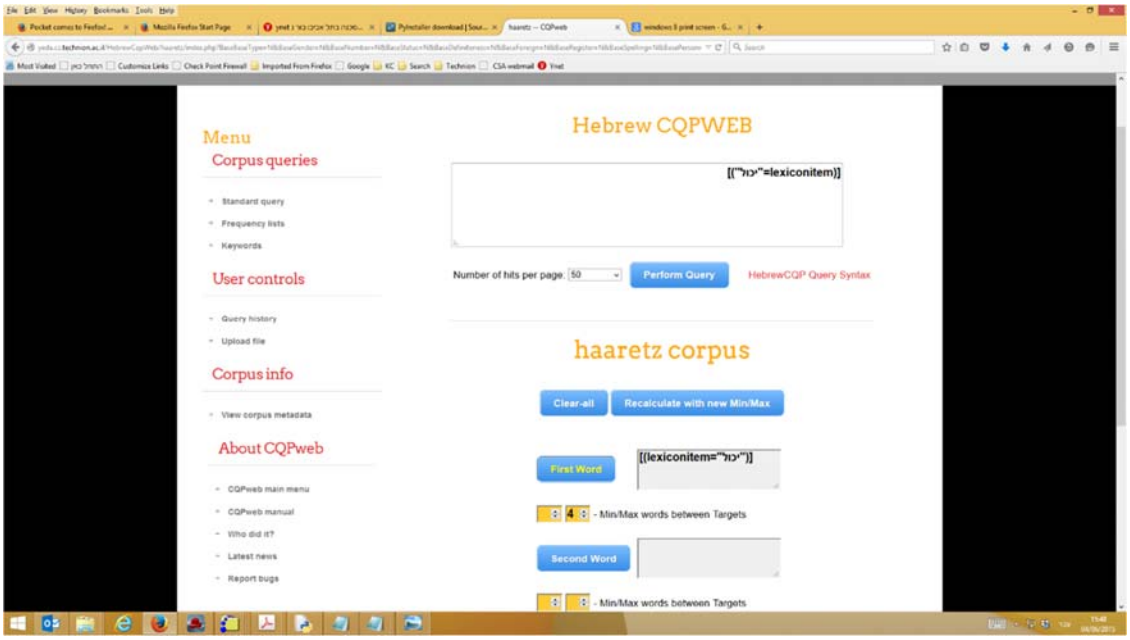


במסך זה סימנו את "יכול" כ-lexical item. הכוונה היא ללמה או לערך המילוני. לחיצה על Submit תוביל אותנו:

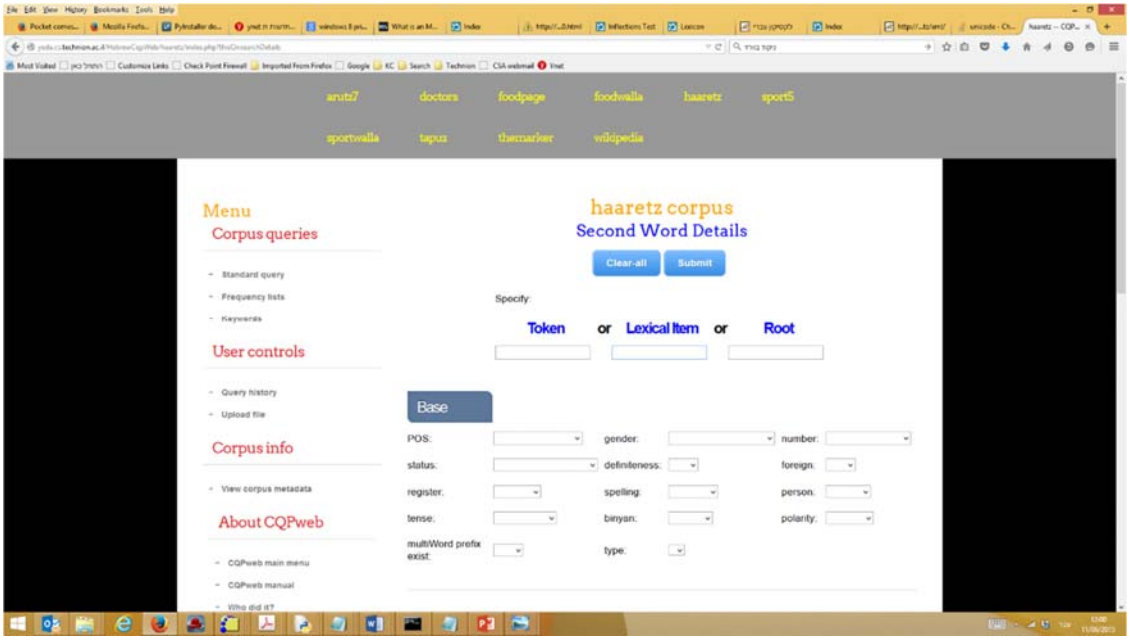


בתיבה העליונה אנו רואים את השאילתה בפורמט של CPQ.

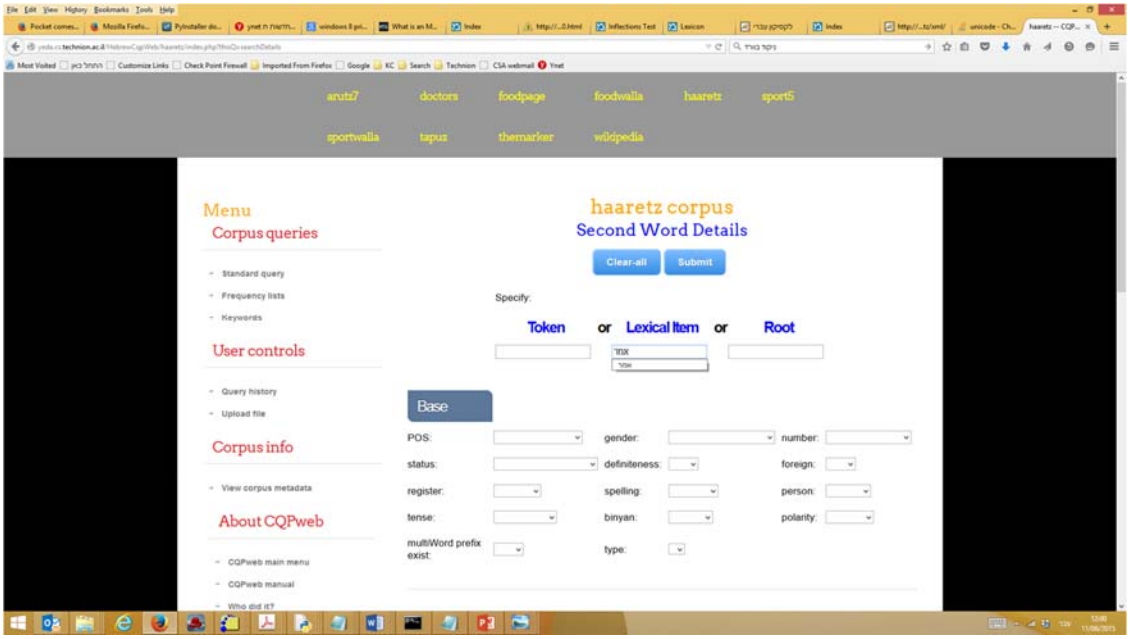
כדי לאפשר למילה השנייה בשאילתה להופיע עד 4 מילים אחרי הראשונה. נציין Min-Max בערך 4 . כאן ציינו את max (אם היה ברצוננו שהרווח יהיה לפחות x מילים היינו משנים את המינימום). כיון שלא שינינו את המינימום אין הגבלה, כלומר המילה השנייה יכולה להופיע מיד אחרי הראשונה, אחרי מילה אחת, ... , אחרי 4 מילים.



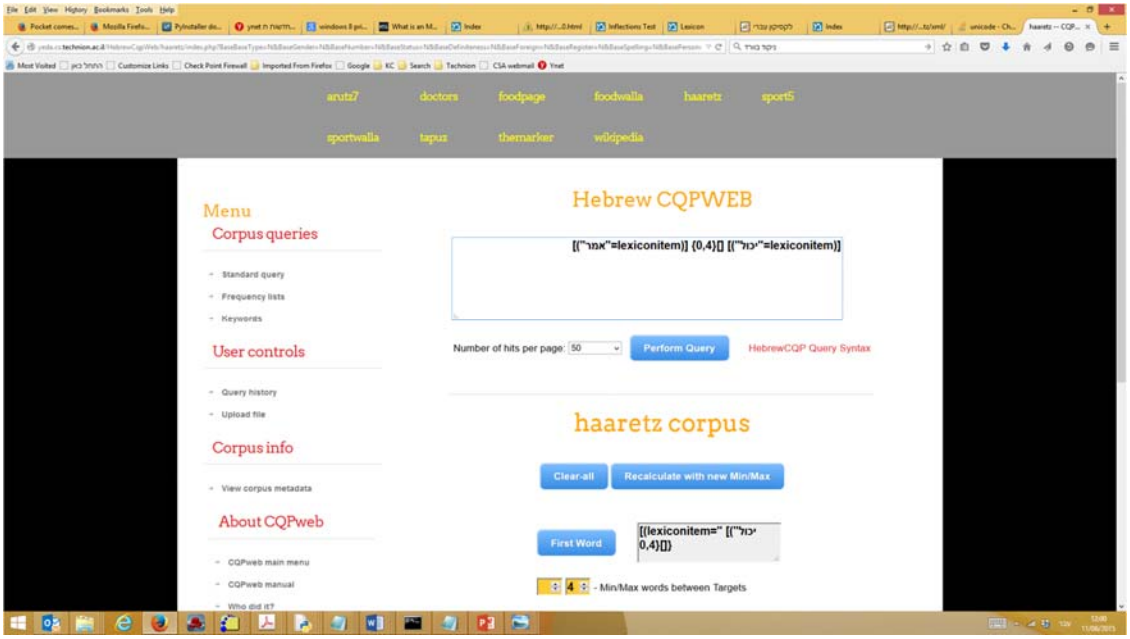
נעבור עתה למילה השנייה (לחיצה על Second Word)



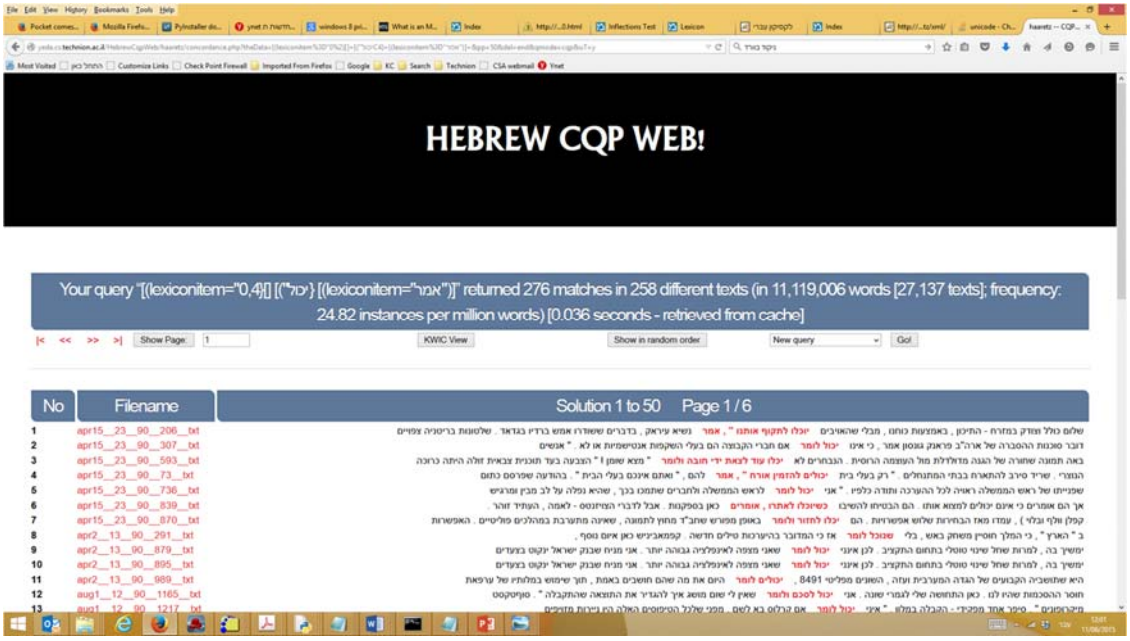
כאן נבחר את הערך הלקסיקלי "אמר"



נלחץ על submit ונראה את השאילתה בפורמט CPQ

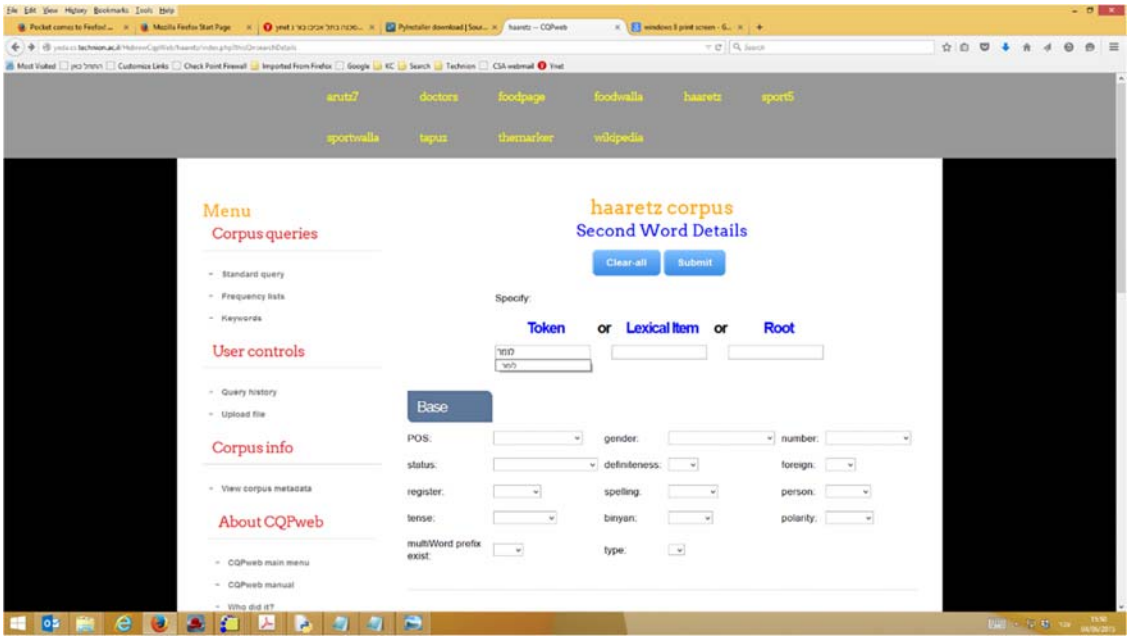


לחיצה על perform query תיתן לנו את תוצאות השאילתה:

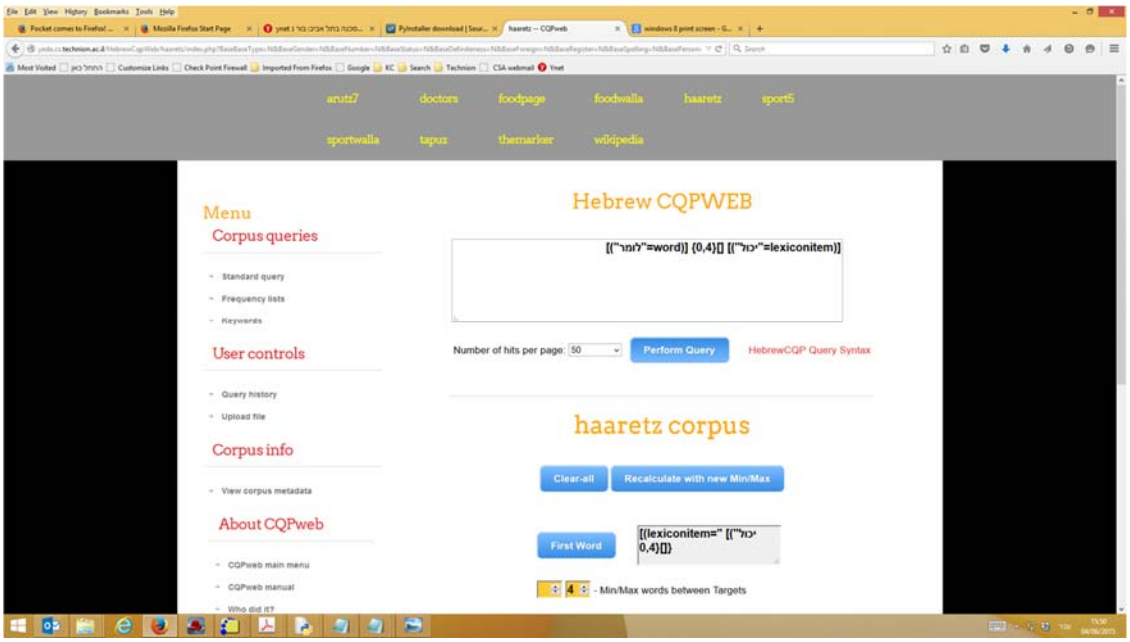


אנו רואים שכל אחת משתי המילים ("יכול" ו"אמר") יכולות להופיע בכל אחת מהנטיות שלהן.

עתה נרצה להגביל את המילה השנייה רק לצורה "לומר". לשם כך נחזור לתפריט המילה השנייה, ננקה אותה (clear) ונרשום "לומר" בתיבה Token.

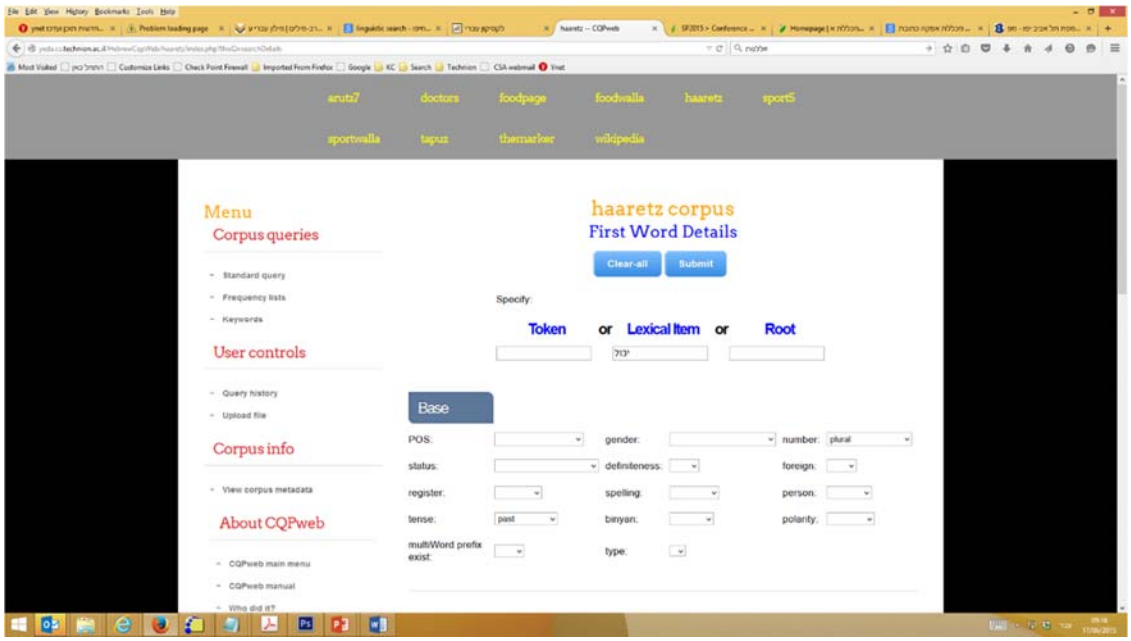


הקלקה על submit תוביל אותנו למסך הבא שבו רואים שהשאילתה מתייחסת למילה השנייה בתור word ולא Lexical Item.

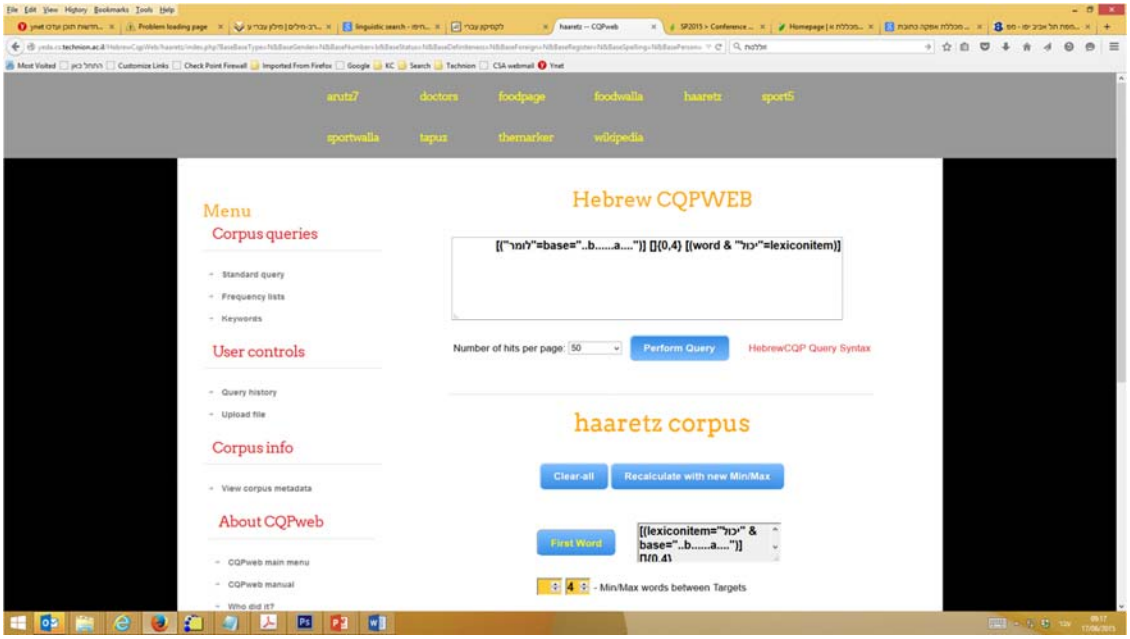


לחיצה על perform query תראה לנו את תוצאות החיפוש. בעוד שהמילה הראשונה מופיעה בנטיות (יכול, יכולים) כל מופעי המילה השנייה הם "לומר".

לבסוף, נגביל את הערכים של המילה הראשונה להטיות של עבר ורבים. לשם כך נחזור לתפריט של המילה הראשונה ונבחר tense = past ו-number = plural.

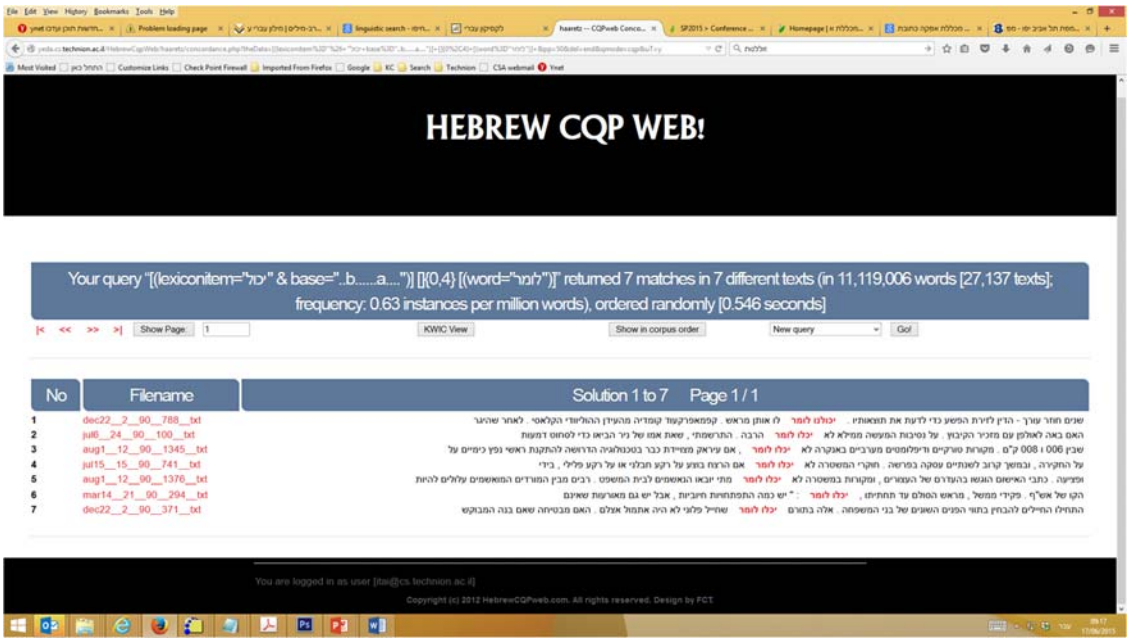


הקלקה על submit תוביל ל-



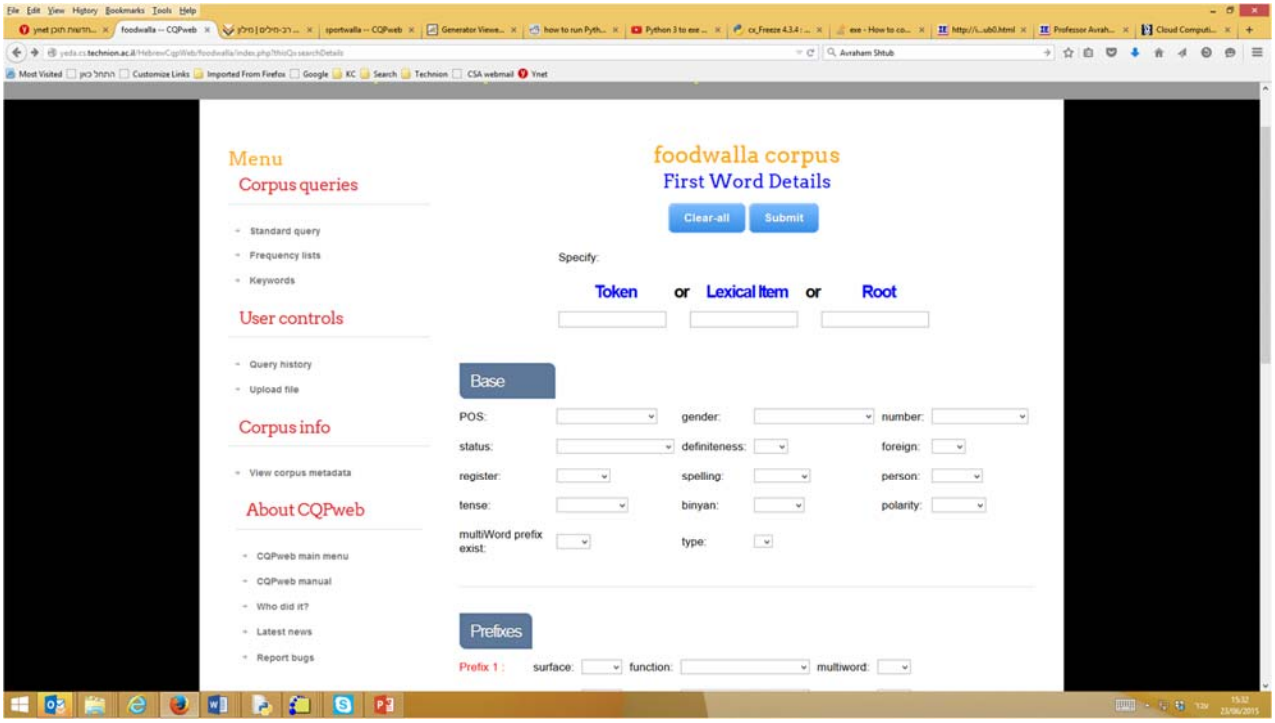
כאן אנו רואים שלמילה הראשונה נוספה דרישת base. (זה לא נראה טוב בגלל שמאל ימין).

הקלקה על perform query מבצע את השאילתה המתוקנת ומראה שבקורפוס הארץ יש 7 משפטים שממלאים את תנאי השאילתה. בצד שמאל מופיעים שמות הקבצים שבהם נמצא כל מופע.

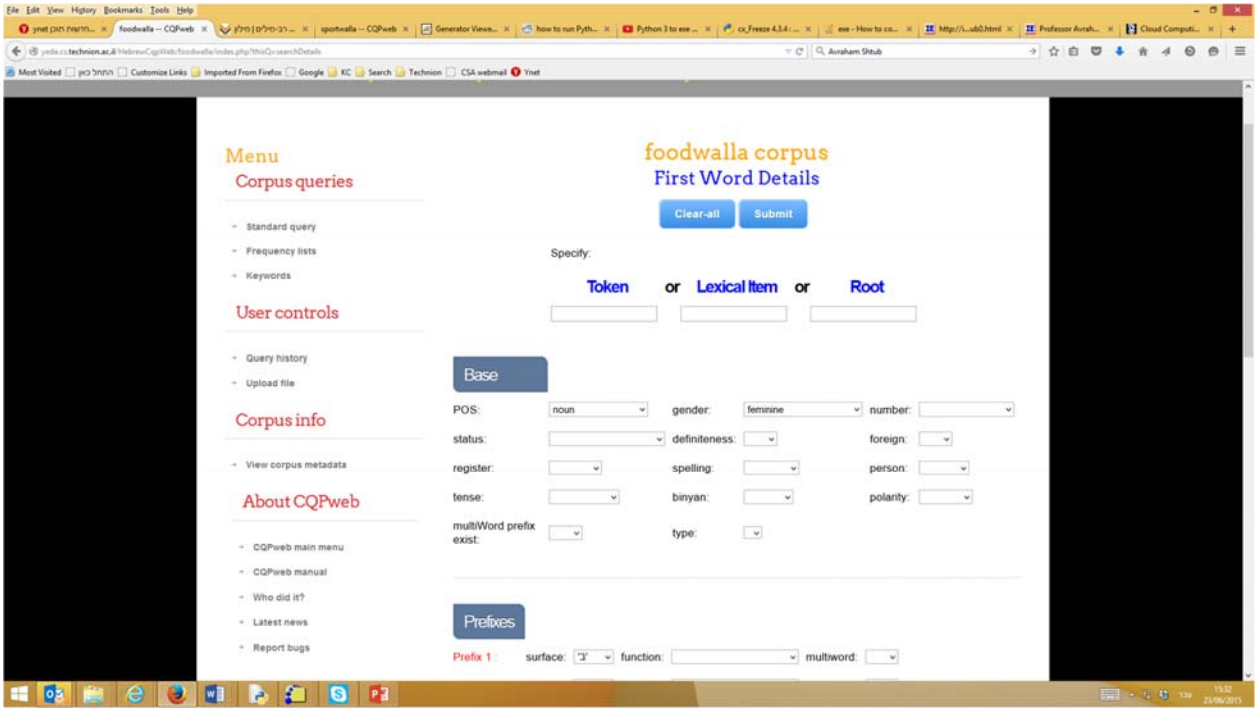


דוגמא 2:

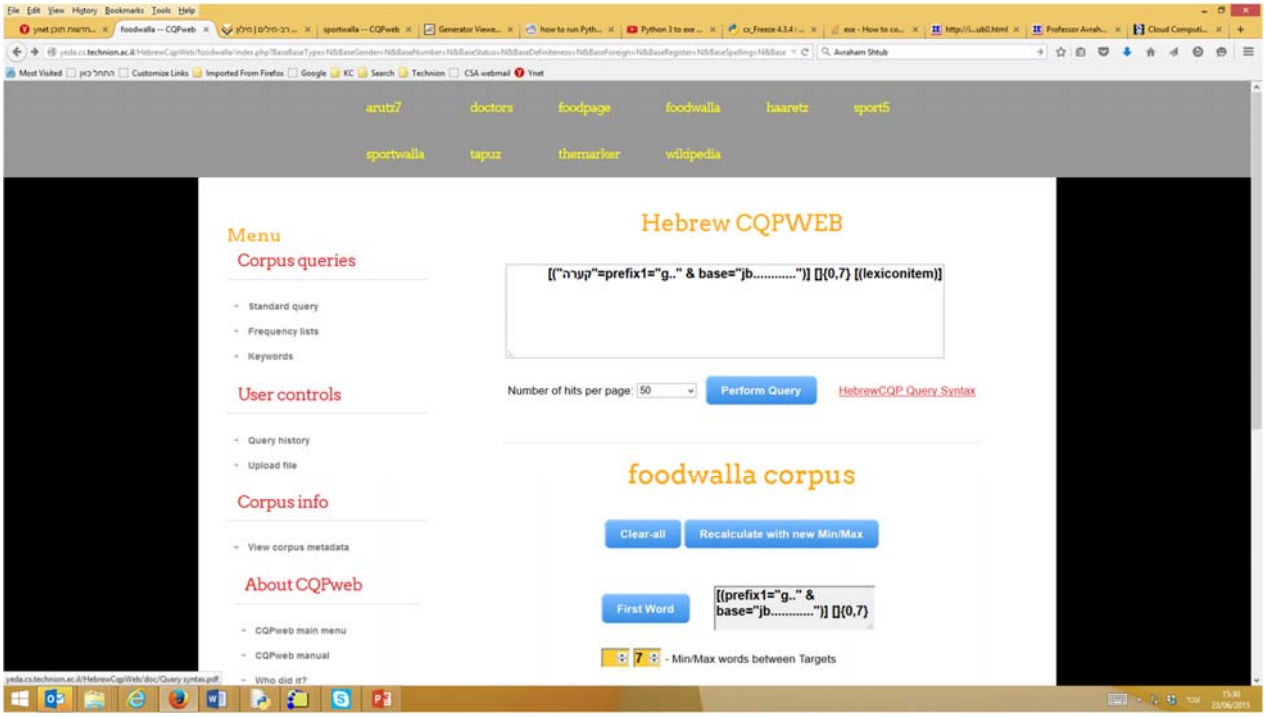
נבחר קורפוס אחר: foodwalla



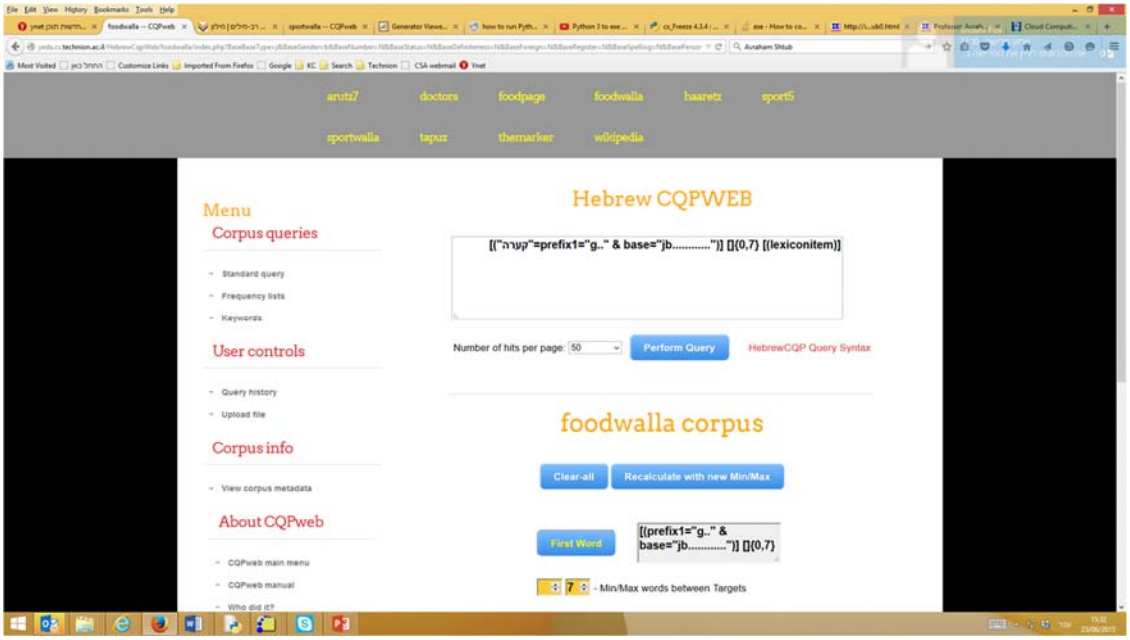
בקורפוס זה, המילה הראשונה תוגבל לשם עצם בנקבה, שמתחילה במילית השימוש "ב":



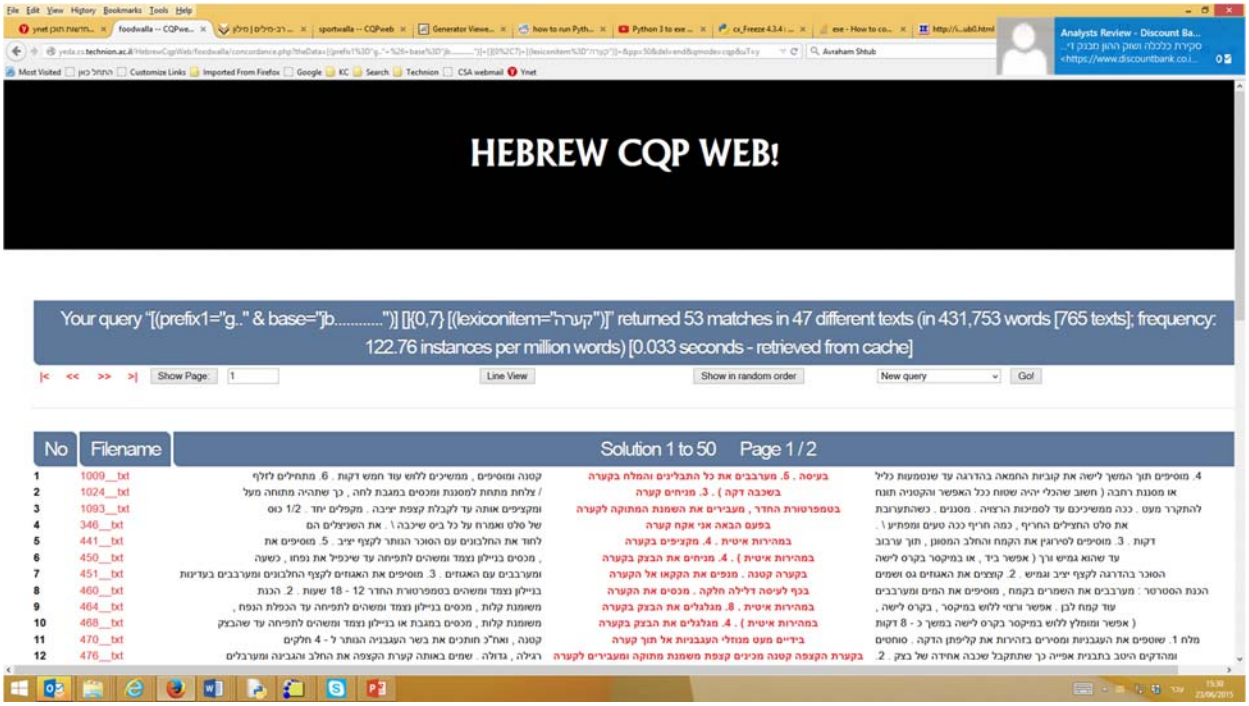
המילה השנייה תבוא אחרי רווח של עד 7 מילים, ותוגבל לערך הלקסיקלי "קערה"



נבצע submit



ולאחריה perform query ונקבל:


HEBREW CQP WEB!
Your query "[(prefix1="g.." & base="]b.....")] []{0,7} [(lexiconitem="קערה")]" returned 53 matches in 47 different texts (in 431,753 words [765 texts]; frequency: 122.76 instances per million words) [0.033 seconds - retrieved from cache]
Show Page: 1
Line View
Show in random order
New query
Go!
No
Filename
Solution 1 to 50 Page 1 / 2
1 1009__txt קטנה ומוסיפים , ממשיכים ללוש עוד חמש דקות . 6. מתחילים לזלף בעיסה . 5. מערבבים את כל התבלינים והמלח בקערה 4. מוסיפים תוך המשך לישה את קובית התמצה בהדרגה עד שנטמעות כליל
2 1024__txt / צלחת מתחת למסננת ומכסים במגבת לחה , כך שתהיה מתוחה מעל בשכבה דקה) . 3. מניחים קערה או מסננת רחבה (חשוב שהכלי יהיה שטוח ככל האפשר והקטניה תוגש
3 1093__txt ומקציפים אותה עד לקבלת קצפת יציבה . מקפלים יחד . 1/2 כוס בטמפרטורת החדר , מעבירים את השמנת המתוקה לקערה להתקרר מעט . ככה ממשיכים עד לסמיכות הרצויה . מסננים . כשהתערובת
4 346__txt של סלט החצילים החריף , כמה חריף ככה טעים ומפתיע \ . בפעם הבאה אני אקח קערה את סלט החצילים החריף , כמה חריף ככה טעים ומפתיע \ .
5 441__txt לחוד את החלבונים עם הסוכר הנותר לקצף יציב . 5. מוסיפים את במהירות איטית . 4. מקציפים בקערה דקות . 3. מוסיפים לסירוגין את הקמח והחלב המסונן , תוך ערבוב
6 450__txt , מכסים בניילון נצמד ומשהים לתפיחה עד שיכפיל את נפחו , כשעה במהירות איטית) . 4. מניחים את הבצק בקערה עד שהוא גמיש וגך (אפשר ביד , או במיקסר בקרס לישה
7 451__txt ומערבבים עם האגוזים . 3. מוסיפים את האגוזים לקצף החלבונים ומערבבים בעדינות בקערה קטנה . מנפים את הקקאו אל הקערה הסוכר בהדרגה לקצף יציב וגמיש . 2. קוצצים את האגוזים גס ושמים
8 460__txt בניילון נצמד ומשהים בטמפרטורת החדר 12 - 18 שעות . 2. הכנת בכף לעיסה דלילה חלקה . מכסים את הקערה הכנת הסטרטר : מערבבים את השמרים בקמח , מוסיפים את המים ומערבבים
9 464__txt משמנת קלות , מכסים בניילון נצמד ומשהים לתפיחה עד הכפלת הנפח , במהירות איטית . 8. מגלגלים את הבצק בקערה עוד קמח לבן . אפשר ורצוי ללוש במיקסר , בקרס לישה ,
10 468__txt משמנת קלות , מכסים במגבת או בניילון נצמד ומשהים לתפיחה עד שהבצק במהירות איטית) . 4. מגלגלים את הבצק בקערה (אפשר ומומלץ ללוש במיקסר בקרס לישה במשך כ - 8 דקות
11 470__txt קטנה , ואח"כ חותכים את בשר העגבנייה הנותר ל - 4 חלקים ביידיים מעט מנוזלי העגבניות אל תוך קערה מלא 1. שוטפים את העגבניות ומסירים בזהירות את קליפתן הדקה . סוחטים
12 476__txt רגילה , גדולה . שמים באותה קערת הקצפה את החלב והגבינה ומערבלים בקערת הקצפה קטנה מכינים קצפת משמנת מתוקה ומעבירים לקערה ומהדקים היטב בתבנית אפייה כך שתתקבל שכבה אחידה של בצק . 2

בתיאבון!