



## המדריך המלא לחיפוש במערכת Buzzilla

יוני 2013

Buzzilla מבצעת בימים אלה את המעבר המלא לסביבת ענן, תוך שדרוג מהותי בכל התשתיות של המערכת. אנחנו עושים את זה גם כי אנחנו רוצים להיות תמיד בקדמת הטכנולוגיה, אך בעיקר כי המעבר יאפשר לנו לספק לכם שירותים עוד יותר מתקדמים, מידע יותר רחב ובעיקר חוויית שימוש אפילו טובה יותר מהמערכת, כפי שאתם מכירים אותה היום.

### היתרונות של המערכת בסביבה החדשה

מערכת Buzzilla המשודרגת תאפשר לנו להתמיד בשיפור המתמיד של חוויית השימוש שלכם. היתרונות המרכזיים הנובעים מהשדרוג הם:

- ◀ שיפור מהותי במהירות החזרת הנתונים. אם עד עכשיו היינו מהירים, עכשיו זה יטוס.
- ◀ עליית מדרגה ברמת הכיסוי. Buzzilla תמיד נתנה את הכיסוי הכי טוב בארץ - עכשיו אנחנו לוקחים את זה עוד כמה מדרגות למעלה. יותר תוצאות, יותר מדויקות, מיותר מקורות.
- ◀ כיסוי גיאוגרפי נרחב. אנחנו מתחילים לכסות את המידע במקומות אחרים בעולם, כמו שאנחנו מכסים אותו בארץ. זה לא יקרה מהיום למחר, אבל בכל מקום שזה יקרה, אתם יכולים לסמוך על כיסוי בסטנדרטים של Buzzilla.
- ◀ פילוח אתרים ופילוח משתמשים עולים מדרגה. מעכשיו התוצאות יכסו את כל השיחות הרלוונטיות לנושא שאתם מחפשים, כך שתוכלו לדעת בצורה, אפילו יותר מדויקת, איפה מדברים וגם מי מדבר.
- ◀ החיפוש נהיה גם יותר פשוט וגם יותר מדויק. החיפוש ב-Buzzilla תמיד היה פתוח, גמיש ופשוט - עכשיו זה נהיה אפילו יותר ידידותי ומאפשר לכם להתמקד בתוצאות רלוונטיות ללא מאמץ.

### השינוי והיתרונות בשיטת החיפוש החדשה

שיטת החיפוש החדשה, עבור השפה העברית, מביאה איתה שתי בשורות נהדרות:

1. השיטה החדשה מגדילה מאוד את רמת הדיוק של תוצאות החיפוש שמתקבלות. המערכת גם תפיק, עבור כל חיפוש, יותר תוצאות, וגם תדאג לכך שהתוצאות יהיו רלוונטיות לנושא שחיפשתם.
2. מעכשיו, הרבה יותר פשוט לכתוב שאילתות "נקיות". תהליך כתיבת השאילתות הופך להיות עוד יותר פשוט. אנחנו עוברים משיטה שבה היה קשה להתמודד עם המורכבות של השפה העברית (הטיות, תחילות, סימומות וכו'), לשיטה, המותאמת במיוחד לשפה העברית, ומאפשרת לנו להתמודד עם המורכבות של השפה ולהניב תוצאות מצוינות.

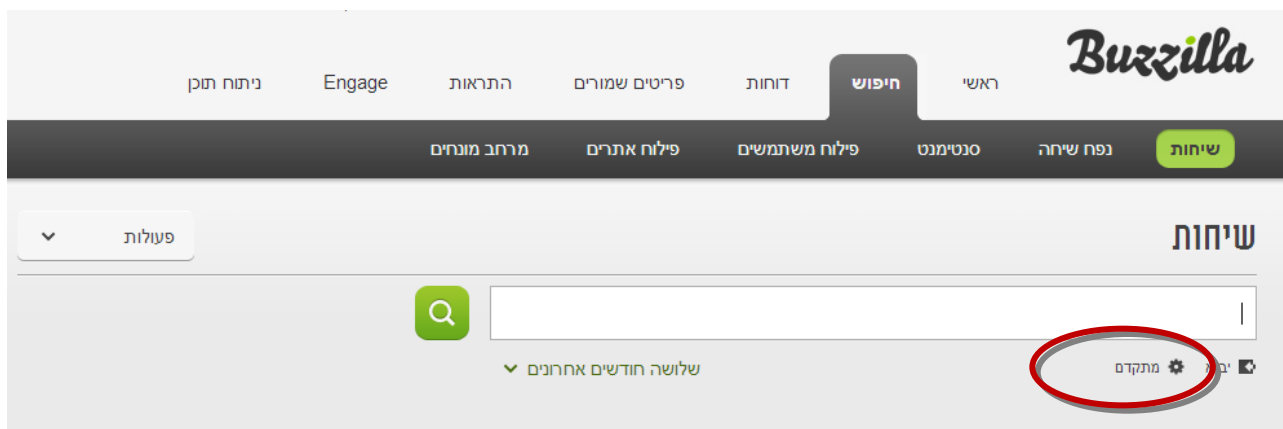
המטרה שלנו בכתיבת שאילתות היא לקבל מהמערכת את אוסף המסמכים (דיונים) המלא והרלוונטי למה שאנחנו מחפשים.

רלוונטיות היא שם המשחק. בחיפוש בכלל ובשפה העברית בפרט, אותה מילה יכולה להיות בעלת משמעות שונה, בהקשרים שונים. למשל, אם נחפש את המילה "דרבי", נקבל תוצאות הקשורות ל-"מכללת דרבי", אך נקבל גם תוצאות הקשורות ל-"דרבי התל-אביבי" (בכדורגל). השאילתא מאפשרת למקד את התוצאות שיתקבלו כך שיהיו רלוונטיות למונחים שאנחנו מחפשים, בהקשר הנכון. אשף השאילתות (חיפוש מתקדם) של Buzzilla מאפשר לבנות שאילתות מורכבות בצורה פשוטה ואינטואיטיבית. השאילתא מורכבת מהרכיבים הבאים:

1. מילים ו/או מונחים (יותר ממילה אחת) המייצגים את הנושא שאנחנו מחפשים
2. מילים ו/או מונחים הממקדים את החיפוש לתחום הספציפי הרלוונטי לחיפוש שלנו.
3. מילים ו/או מונחים שאנחנו לא מעוניינים למצוא מסמכים שבהם הם מופיעים.
4. פילטרים (זירות, אתרים, שפות, פרמטרים של מעורבות וכו').

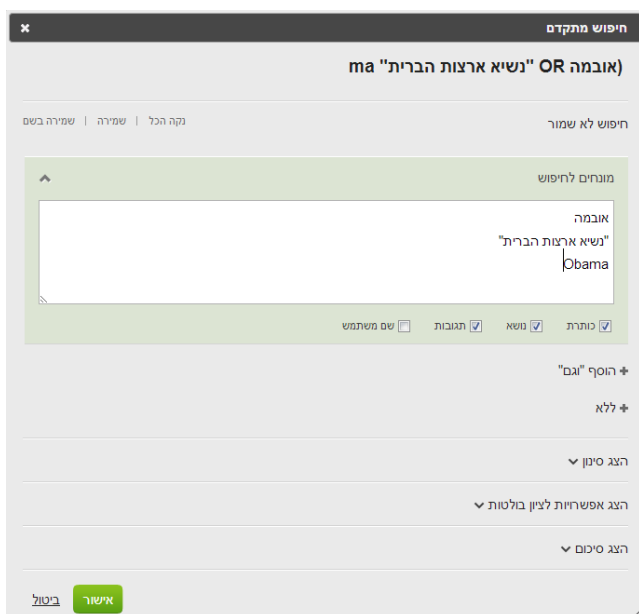
שאילתא תמיד תכיל את המילים/מונחים המייצגים את הנושא שאנחנו מחפשים. את שאר רכיבי השאילתא ניתן לכלול לפי הצורך.

על מנת לכתוב שאילתא במערכת נפתח את אשף השאלות על ידי הקלקה על "מתקדם" (מופיע מתחת לשדה מונחים, בצד ימין, בכל מקום במערכת בו מופיע שדה מונחים)



## הזנת מונחים בסיסית

מונח יכול לכלול מילה בודדת, לדוגמא: **אובמה**, או יותר ממילה אחת, לדוגמא: **נשיא ארצות הברית**. בחלון "מונחים לחיפוש" מקלידים את המונחים המתארים את הישות או הנושא אותו אנחנו מחפשים.



- ◀ אם המונח שלנו הינו מילה בודדה, פשוט נכתוב אותה בשדה ה-"מונחים לחיפוש".
- ◀ אם המונח שלנו כולל יותר ממילה אחת - נכתוב את כל המונח במירכאות.
- ◀ אם יש יותר ממונח אחד המשמש להתייחסות לנשוא החיפוש שלנו, נכניס מונח אחד בשורה, נשבור שורה ונכניס את המונח הבא. (שיטת כתיבה זו מורה למערכת להחזיר את כל הדיונים שבהם מופיע **לפחות אחד** מהמונחים שהוכנסו).
- ◀ ניתן להכניס בשדה זה מונחים בשפות שונות
- ◀ ניתן לכתוב מילים שיש בהן גרשיים, לדוגמא: **רמטכ"ל** וגם מילים שיש בהן גרש, לדוגמא: **אורנג'**.

### חיפוש "וגם"

ניתן להשתמש באפשרות החיתוך "וגם", כדי לקבל תוצאות המכילות צירוף של מונחי החיפוש שאותם הגדרת בחיתוך. לשם כך לחץ על + הוסף "וגם" לפתיחת חלונית זו. גם בתוך שדה החיתוך ניתן להגדיר מספר מונחים שהיחסים ביניהם הם "או" (בשורה נפרדת). בדוגמא שלפנינו, ברצוננו לנטר את השיחות על עישון בהקשר של הגבלת העישון במקומות ציבוריים, ולכן נוסיף בשדה החיתוך "וגם" את המילים: "פאב, בית קפה, מסעדה, מקום ציבורי"

חיפוש מתקדם: חיפוש "וגם"

חיפוש זה יחזיר את כל התוצאות שבהן תופיע לפחות אחת מהמילים "עישון", או "מעשן" וגם "פאב" או "בית קפה" או "מסעדה" או "מקום ציבורי" (לפחות אחד מהמונחים בכל אחד מהשדות או יותר).

### החרגת מונחים ("ללא")

בשדה זה ניתן להגדיר מונחים שאותם רוצים להוציא לגמרי מתוצאות החיפוש. למשל:

חיפוש מתקדם: החרגת מונחים

יחזיר רק שיחות שבהן מוזכר המונח "עישון" שלא בהקשר למזונות המעושנים ("בשר" ו-"נקניק").

## צמצום אזורים בשיחה

ניתן למקד את החיפוש ע"י ביטול סימון התיבות מתחת לשדה, המאפשרות לקבוע אם החיפוש יכלול את המופע של מונחי החיפוש בכותרת, בנושא או בתגובות לשיחה, או בשמות המשתמשים המשתתפים בשיחה. בחירה או ביטול של מיקום מונחי החיפוש זמינה בכל החלוניות הדומות בחלון "חיפוש מתקדם". אם מעניין אותך למצוא שיחות בנושא עישון, אך לא את אלה שבהן נושא העישון הוזכר בהן רק בתגובות לכתבה, לפוסט או להודעה הראשית, ניתן להגביל את החיפוש לכותרת ולנושא בלבד, באמצעות ביטול הסימון ליד "תגובות".



חיפוש מתקדם: ביטול החיפוש בתגובות

מערכת Buzzilla מחלקת את השיחה לארבעה אזורים:

- ◀ **כותרת:** כותרת של הכתבה, הפוסט או ההודעה הראשית בשרשור של פורום (ללא כותרות משנה)
- ◀ **נושא:** תוכן הכתבה, הפוסט או ההודעה הראשית בשרשור של פורום
- ◀ **תגובות:** התגובות לכתבה, לפוסט, לסטטוס (בפייסבוק, למשל) או לציוץ (בטוויטר), וכן הודעות התגובה בשרשור של פורום.
- ◀ **שם משתמש (אינו מסומן כברירת מחדל):** סימון תיבה זו יחזיר גם תוצאות שנמצאו בשדה שם המשתמש. עם זאת, בשדה זה יוחזרו רק מסמכים שיש בהם התאמה מלאה בין המונח לבין שם המשתמש שנסרק. למשל: חיפוש שם המשתמש "אחד שידוע" יביא רק שיחות שבהן השתתף משתמש בשם "אחד שידוע", אבל לא יאותרו שיחות שבהם השתתף המשתמש "אחד שידוע יותר", למשל.

בלחיצה על הצג סינון נפתחות מספר אפשרויות נוספות לצמצום החיפוש לשם הגברת דיוק התוצאות:

חיפוש מתקדם: אפשרויות סינון מתקדמות

### 1. סינון לפי מאפייני אתר

- ◀ **סינון זירות:** כברירת מחדל, Buzzilla מחפשת שיחות מכל הזירות: בלוגים, טוויטר, כתבות, פורומים ורשתות חברתיות. כדי לצמצם את החיפוש ולמקד אותו בסוג מסוים של שיחות, סמן את האפשרות "הצג רק מזירות אלו", ולאחר מכן הסר את הסימונים מסוגי הזירות שלא מעניינים אותך.
- ◀ **סינון אתרים:** מתבצע בשתי צורות. ניתן להגדיר רשימה של אתרים שהחיפוש יתמקד בהם (לבחור "להציג רק מאתרים"), או רשימה של אתרים שלא יכללו בחיפוש (לבחור "לא להציג מהאתרים" – צבע הרקע של השדה משתנה לאדום). בשדה זה יש לכתוב בכל שורה שם של אתר אחד, כולל סיומות כתובת ותחום במירכאות כפולות, לדוגמא: "tapuz.co.il".

### 2. סינון לפי מאפייני השיחה

- ◀ **לפי שפה:** כברירת מחדל, Buzzilla מחפשת התאמות בשיחות מכל השפות. ניתן להגביל את החיפוש לשפה מסוימת. סינון לפי שפה שימושי במיוחד כשמעוניינים לאתר את השיח סביב נושא המנוסח בשפה אחרת. למשל, אם אתה חוקר את השיח סביב המותגים CNN, Nokia או Microsoft, אך מעוניין להתמקד בשיח העברי בלבד.

◀ **לפי מאפייני חשיפה:** קיימים כמה מדדים מספריים המעידים על מידת העניין שעוררה שיחה מסוימת ברשת, או על מידת החשיפה שקיבלה. המערכת מאפשרת לקבוע רף מינימום של מדדים אלה בתוצאות החיפוש המוחזרות:

- **תגובות:** מספר התגובות שקיבל הדיון
- **משתתפים:** הנתון אינו בהכרח זהה למספר התגובות שכן לעיתים משתתף אחד כותב מספר תגובות בתוך דיון בודד.
- **הצבעות:** נמדדות רק בפייסבוק (מספר הלייקים) וב-Google Plus (מספר ה-1+ שקיבל הסטטוס).
- **שיתופים:** נמדדים רק בפייסבוק (מספר ה-Share) ובטוויטר (מספר ה-ReTweets).

### 3. סינון לפי מאפיינים ייחודיים לפייסבוק

סוג הסטטוס: סטטוס רגיל (טקסט בלבד), וידאו, תמונה או קישור.  
סוג המחבר: מי כתב את זה? ניתן לסנן לפי "עמוד אוהדים" (Page) או "פרופיל אישי" (Profile)

## שליטה על הטיות ותחיליות

השפה העברית הינה שפה מורכבת במונחים של מנועי חיפוש. לעיתים אותה מילה יכולה לקבל שתי משמעויות, לדוגמא: **בזק** (החברה) ו-**בזק** (כמו "מהירות בזק"). לעיתים קרובות מילים בעברית הנן כהטיות של אותם שורשים, כמו למשל **בזק** ו-**מבזק**. במקרים רבים אותה מילה יכולה לקבל אותיות תחיליות וסימומות (כמו "ה", "ו", "בכל" מ וכו'), לדוגמא: **מבזק** ו-**שבמבזק**.

בעבודה עם שאילתות, לעיתים אנחנו רוצים שהמערכת "תנחש" הטיות רלוונטיות של המונחים שאנחנו מחפשים ותחזיר לנו תוצאות גם עבור **מורה** וגם עבור **מורים**, אך לעיתים אנחנו רוצים רק את המונח הספציפי או רק אם הוא מופיע עם חלק מהתחיליות. לדוגמא: כשאנחנו מחפשים את המונח "**בזק בינלאומי**", נרצה לקבל תוצאות שבהם הוזכר המונח "**ובזק בינלאומי**" אבל לא נרצה לקבל תוצאות שבהן הוזכר המונח "**הבזק בינלאומי**". מנוע החיפוש החדש של Buzzilla מאפשר דיוק מרבי בהגדרה של המונחים הרצויים, זאת בעזרת מספר מנגנונים להגדלת הדיוק:

### 1. מילים מוכרות ו-"לא מוכרות"

מנוע החיפוש של Buzzilla מושתת על מילון המזהה מילים, גם בעברית וגם כאלה שאינן בעברית (כמו שמות מותגים) אך נכתבות בעברית ולרלוונטיות לקהל דובר עברית (לדוגמא: **מיקרוסופט**). נוכל לדעת אם מילה שהכנסנו מזהה על ידי המילון או לא שכן מילים שאינן מזהות על ידי המילון יופיעו עם קווקוו אדום מתחתם לדוגמא: **ציפרלקס**. מילים מזהות לא יקבלו שום סימון. כשאנחנו מכניסים לשדה המונחים מילה המזהה על ידי המילון, ברירת המחדל של המערכת תהיה להחזיר לנו תוצאות עבור כל ההטיות של המונח ועבור כל צירופי התחיליות והסימומות. כך, כשנחפש את המילה **מחשב** נקבל גם תוצאות שבהן הופיעו המונחים **מחשבים**, **חשב**, **חישב** (לרבות הטיות שלהם והופעותיהם



עם תחיליות וסיומות). נוכל לקבל חיווי אם המילון "משייך" את המילה שכתבנו למילים אחרות - אם נעביר את הסמן מעל המילה שברצוננו לבדוק, נקבל חלונית שבתוכה מצוינות המילים הנוספות המשויכות למילה שלנו. כך נוכל להחליט אם ברצוננו לקבל תוצאות עבור כל ההטיות האפשריות של המילה או שברצוננו להגביל להטיות ספציפיות בלבד.



על הדרך להגבלת ההטיות והתחיליות נרחיב להלן בסעיף "הגבלת הטיות ותחיליות". עבור מילים "לא מוכרות", ברירת המחדל של המערכת תהיה, להחזיר תוצאות שבהן מופיע המונח המדויק, ללא כל הטיות, תחיליות או סיומות. לדוגמא: אם נחפש **ציפרלקס**, לא נקבל תוצאות שבהן הופיע המונח **וציפרלקס**.

## 2. מונחים שיש בהם יותר ממילה אחת (ומורכבים ממילים מוכרות)

בכתיבת מונחים המורכבים ממילים מזוהות, ברירת המחדל תהיה שהמערכת תחזיר תוצאות הכוללות את המונח המלא, או כל מונח המורכב מהטיות של המילים המרכיבות את המונח המקורי. לדוגמא: נחפש את המונח "**שלומי שבת**". אם נעביר את הסמן על המילה **שלומי**, נראה שמילה זו משויכת גם למילה **שלום**. לכן, בעבור חיפוש זה, המערכת תחזיר גם תוצאות שבהן הופיע המונח "**שלום שבת**" וגם תוצאות שבהן הופיע המונח "**שבת שלום**".



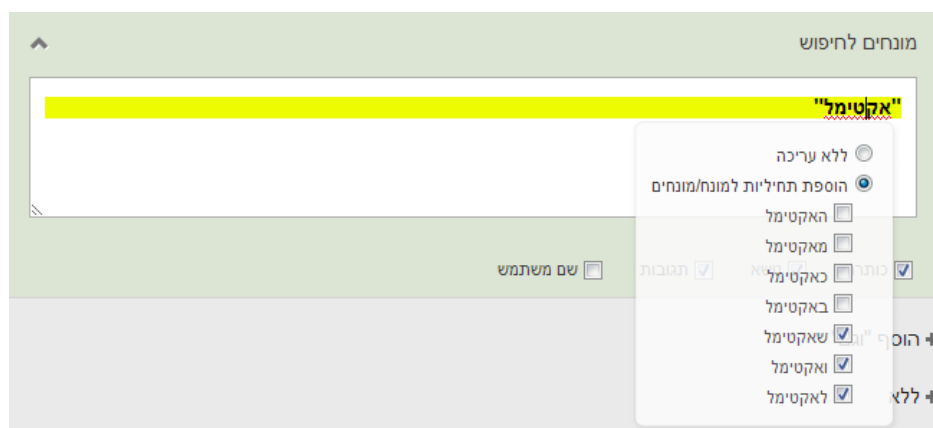
נוכל להגביל את ההטיות שעבורן יוחזרו תוצאות על ידי שימוש בהגבלת הטיות ותחיליות כפי שיפורט בסעיף הבא.

יש מקרים שבהם נצליח לדייק את התוצאות שנקבל עבור החיפוש אם נבטל את הטייות למונחים שלנו ונגדיר אילו תחיליות רלוונטיות לנו. ממשק החיפוש המתקדם מאפשר להגדיר הגדרות אלה בצורה פשוטה. ההגדרה תתבצע עבור כל שורה, בחלון המונחים לחיפוש, בנפרד. ההגדרות הן עבור השורה המלאה, גם אם יש בה מילה אחת וגם אם יש בה מונח שמורכב מיותר ממילה אחת.

### מקרה 1: הגבלת הטייות ותחיליות עבור מונח שאינו מזהה על ידי המערכת

אם נכניס את המונח **אקטימל** לחלון המונחים לחיפוש, נראה שהמונח מקבל קווקוו אדום שמשמעו שהמונח אינו מזהה על ידי המערכת. המשמעות היא שאם לא נעשה כלום, השאילתא הזאת תחזיר רק תוצאות שבהן מופיע המונח המדויק. לא יופיעו תוצאות שבהן הופיע המונח עם אות תחילית, לדוגמא: **ואקטימל**. על מנת להגדיר שהחיפוש יניב תוצאות גם עם תחיליות, נקליק עם העכבר **הימני** על המילה **אקטימל**. יפתח תפריט צף שבו שתי אפשרויות לבחירה:

1. "ללא עריכה" - זאת האפשרות המסומנת כברירת מחדל.
2. "הוספת תחיליות למונח/מונחים" - נסמן את הכפתור שליד אפשרות זו.



אפשרות זו מאפשרת לנו לבחור אילו תחיליות רלוונטיות למונח שלנו. נוכל עכשיו לסמן ליד כל מילה  אם אנחנו רוצים לקבל אותה בתוצאות או להשאיר ריבוע סימון ריק אם לא. שימו לב ששלוש האפשרויות התחתונות מסומנות מראש, שכן אלו תחיליות המתאימות כמעט לכל מונח. ניתן להסיר מהן את הסימון אם הן לא רלוונטיות. המערכת תוסיף באופן אוטומטי את כל צירופי התחיליות הנפוצים הרלוונטיים לאופציות שסימננו, לדוגמא: אם סימננו את האפשרות **האקטימל**, המערכת תגדיר גם את **הואקטימל** כתוצאה רצויה. לאחר שסיימנו לסמן, נקליק הקלקה רגילה באזור אחר של המסך על מנת לסגור את התפריט הצף. המונח **אקטימל** יופיע עכשיו בחלון מונחי החיפוש כשהוא מודגש בבולד ותחום במירכאות. כך נדע לזהות בכל שאילתא, אילו מונחים קיבלו עריכה ואילו לא.

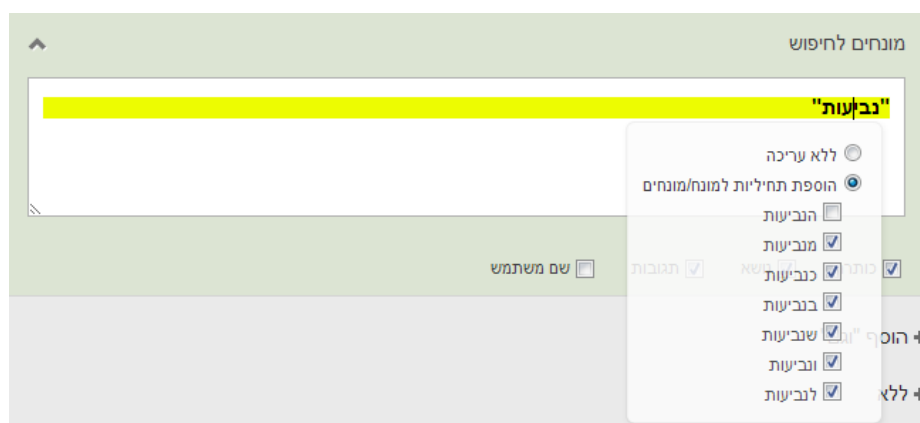


## מקרה 2: הגבלת הטיות ותחיליות עבור מונח מזהה

אם נכניס את המונח **נביעות** לחלון המונחים, המונח לא יקבל כל סימון שכן זוהי מילה מזהה. אם נעביר את הסמן מעל המילה נביעות (ללא לחיצה), המערכת תראה לנו שמילה זו משויכת להטיה **נביעה**. משמעות הדבר היא שאם לא נבצע עריכה, שאלתא אז תניב גם תוצאות שבהן נמצאה המילה נביעה (והטיותיה השונות).



אם ברצוננו לחפש דיונים אודות המונח **נביעות** נצטרך לבטל את ההטיות ולהגביל את התחיליות. נעשה זאת על ידי הקלקה ימנית על המילה נביעות, וסימון האפשרויות המתאימות בתפריט הצף שיפתח (כמוסבר ב-"מקרה 1").



שימו לב שלא סימננו את האפשרות **הנביעות** מתוך הנחה שהצירוף הזה לא מתאים להתייחסויות למותג ולכן יניב תוצאות לא רלוונטיות.

### מקרה 3: הגבלת הטייות ותחיליות עבור מונח שיש בו יותר ממילה אחת

אם נכניס את המונח **מכללת אפקה**, כשבכוונתנו למצוא אזכורים של המכללה. אם נעביר את הסמן על המילה **מכללת** נגלה שמילה זו משויכת למילה **מכללה**. אם נעביר את הסמן על המילה **אפקה**, נגלה שמילה זו משויכת גם למילים **אפיק** ו-**אופק**. אם לא נבצע עריכה יתקבלו תוצאות המכילות צמדי מילים שהן הטייות של מילים אלה, לדוגמא: **המכללה באופק**.

הפעלת העריכה על השורה הזאת תשאיר את המונח כמו שהוא ותאפשר לקבל אותו עם התחיליות הרלוונטיות שנסמן. לצורך הפעלת העריכה, נקליק עם העכבר הימני על אחת המילים בשורה - יקפוץ תפריט העריכה הצף. כאן נסמן את האפשרות השניה ונבחר את האפשרויות הרלוונטיות:

