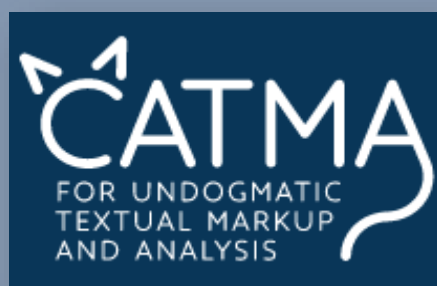


מחקר טקסטואלי חישובי במערכת CATMA

(Computer Assisted Textual Markup & Analysis)



מדריך עברי

מהדורה נסיונית ראשונה | יוני 2017

איתי מרינברג-מיליקובסקי

למה CATMA?

[1. מבוא עמ' 4](#)

לפני שמתחילים

[2. לפני שמתחילים \(א\): התקנת קורפוס לדוגמא עמ' 6](#)

[3. לפני שמתחילים \(ב\): אופרציונליזציה ראשונית עמ' 8](#)

צעדים ראשונים

[4. רישום וכניסה למערכת עמ' 12](#)

[5. רכיבי המערכת: מבט כללי עמ' 14](#)

[6. הזנת טקסט, יצירת קורפוס עמ' 17](#)

ניתוח אוטומטי, טיפול ב- data

[7. ניתוח אוטומטי ראשוני של טקסט או של קורפוס שלא תויגו עמ' 20](#)

[8. בונה השאלות: היכרות ראשונית \(שאלות פשוטות\) עמ' 26](#)

[9. בונה השאלות: שאלות מורכבות עמ' 34](#)

תיוג ידני, יצירת metadata

- [10. תיוג פרשני: מבוא קונספטואלי](#) עמ' 41
- [11. גיבוש מערכת תגים](#) עמ' 43
- [12. הכנת מסך התיוג \(Tagger\): טעינת טקסט, מערכת תגים ואסופת תיוג](#) עמ' 51
- [13. תיוג ידני – הלכה למעשה](#) עמ' 58
- [14. מתיוג ידני לניתוח אוטומטי: תיוג אוטומטי-למחצה](#) עמ' 68

לקראת מחקר: מבט חישובי על תובנות אנושיות, ולהיפך

- [15. ניתוח אוטומטי של טקסט או של קורפוס מתויג \(א\): שאילתות פשוטות](#) עמ' 73
- [16. ניתוח אוטומטי של טקסט או של קורפוס מתויג \(ב\): בניית לקסיקון מבוסס-תיוג](#) עמ' 80
- [17. ניתוח אוטומטי של טקסט או של קורפוס מתויג \(ג\): שאילתות מורכבות](#) עמ' 82

מבט נוסף

- [18. ניתוח חוזר: מהכניסה השניה למערכת ואילך](#) עמ' 88
- [19. ייצוא נתונים וממצאים, שיתוף מידע](#) עמ' 90

אחרית דבר

- [20. במקום סיכום: עקרונות מנחים לעבודה במערכת CATMA](#) עמ' 98
- [21. המלצות לקריאה נוספת](#) עמ' 100
- [22. נספח: שפת השאילתות](#) עמ' 101

1.1. מערכת CATMA (Computer Assisted Textual Markup and Analysis) נועדה לתמוך במחקר חישובי של טקסטים. היא מותאמת לכל השפות (כמעט) – לרבות עברית – ומאפשרת גמישות מלאה בבחירת יעדי המחקר. היא אינה מוגבלת לדיסציפלינה מסוימת, ולפיכך עשויה לתרום לכל מחקר טקסטואלי במדעי הרוח והחברה. השימוש בה אינו כרוך בתשלום ואינו מחייב ידע מוקדם בתכנות או במדעי המחשב. המערכת היא פרי פיתוח נמשך והולך של צוות חוקרים מן המרכז האינטרדיסציפלינארי לנראטולוגיה (Interdisciplinary Center for Narratology) באוניברסיטת המבורג, גרמניה, בראשות פרופ' יאן כריסטוף מייסטר.

1.2. CATMA חותרת לשלב יחד שתי צורות ניתוח: מחד, ניתוח אוטומטי ממוחשב של הטקסט, המבוסס על מיפוי לקסיקלי, על איתור דפוסים לשוניים חוזרים, וכיוצא באלה תכונות 'אובייקטיביות'; מאידך, תיוג או ביאור (annotation) ידני חופשי של הטקסט באמצעות מערכת סימון משוכללת, המקפלת בתוכה יסודות פרשניים 'סובייקטיביים' שהוגדרו בידי המשתמש. בעשותה כן, כמו גם בפנותה לחוקר ה'מסורתי' שלא חונך על ברכי מדעי הרוח הדיגיטליים, סוללת המערכת נתיב ייחודי בתחום: הנחות היסוד עמן ניגש החוקר אל הטקסט, תובנותיו, מסקנותיו, אפילו שאלותיו ותהיותיו – כל אלה אינן נדחות בהכרח בפני נתונים דיגיטליים נוקשים שהופקו בעזרת מכונה חסרת בינה, ואף לא מוכנסות לסד נוקשה מדי של קטגוריות מוכנות-מראש, אלא הופכות לשכבת מידע יצירתית וגמישה על גבי הטקסט. החוקר יכול לעבור לסירוגין מצורת ניתוח אחת אל השנייה, להצליב ביניהן, ואף לעמת את הממצאים העולים מהן. שתי שכבות המידע – האוטומטית והידנית – משלימות זו את זו ומאתגרות זו את זו, בסביבה המעודדת ניתוח שיטתי, מחוייב וקפדני; הן מספקות תשתית אופטימלית לקריאה ביקורתית רפלקסיבית, מודעת לכוחה ולמגבלותיה, נכונה לשפר את עצמה בכל עת.

1.3. CATMA עובדת היטב עם טקסטים בכל גודל והיקף (כמעט), מן הפרגמנט הקצר הבודד, אל הקורפוס העצום, מרובה-הטקסטים. נאמנה לשיטתה, היא מאפשרת תנועת רצוא ושוב בין קריאה קרובה (close reading) וקריאה רחוקה (distant reading); ובהיותה מערכת מקוונת, תומכת המערכת לא רק בעבודת החוקר היחיד, אלא גם בקריאה ובניתוח שיתופיים, בכל שלבי העבודה.

1.4. מדעי הרוח הדיגיטליים – בעיקר בגילומם כאסטרטגיה פרשנית (להבדיל מן הדימוי הרווח שלהם כמארג כלים

לדיגיטציה של חומרים לא-דיגיטליים) – מעוררים לעתים חשש ואף התנגדות בקרב הומניסטים 'מסורתיים', מסיבות שלא כאן המקום לעמוד עליהן. המדריך המוצע כאן מתמקד בסביבה חישובית שלגביה אין לחששות כאלה מקום: היא מאמצת את הפרקטיקות ההרמנויטיות הישנות והטובות, ורק מוסיפה להן כח ועוצמה. בה בעת, כפי שייווכח לדעת כל מי שישתמש במערכת לאורך זמן, אי אפשר לומר שאין בה חידוש פרדיגמטי כלשהו.

1.5. המדריך שלהלן מתייחס לגרסה 5.0 של המערכת בצורתה העדכנית ביותר (נכון לזמן כתיבת שורות אלה). מטרתו

אינה רק להציג את CATMA בפני חוקרים דוברי עברית ולעמוד על מקצת ההיבטים המייחדים ניתוח של טקסטים בעברית, אלא גם ללוות את החוקרים עקב בצד אגודל בשלבי ההתנסות הראשוניים במערכת; אפשר לראותו כמעין סדנה בכתב, וכדאי להקדיש לו שעות אחדות. המדריך אף עוקב אחר פיתוחה של דוגמת-מחקר בזעיר אנפין – שני טקסטים קצרים ושאלת מחקר אחת – שסביבה נעות רבות מן ההדגמות בתהליך הלימוד. המשתמש, כפי שיצוין בהמשך, מוזמן לגשת ללימוד המערכת תוך התנסות ממשית בטקסטים המעסיקים אותו.

1.6. לאחר התנסות קצרה המערכת עתידה להתגלות כאינטואיטיבית להפליא. ברם, ההגעה לשלב זה דורשת לימוד

ותרגול חוזר. כך למשל אחד מעקרונות היסוד של CATMA – עקרון הפרדת השכבות, שעוד ידובר בו רבות – עלול להיראות תחילה כגורם מסרב ומעכב, ונחוצה סבלנות כלשהי כדי לעמוד על הפוטנציאל המחקרי הגלום בו. לשון המדריך, לפיכך, אינה מינימליסטית או טכנית; מאמץ מיוחד הושקע על מנת להתמודד עם שגיאות נפוצות בהבנת המערכת ובעבודה עימה, וצילומי מסך מפורטים צורפו כמעט לכל שלב. הוא קשוב במיוחד לסוגיות המעסיקות חוקרים שאינם רגילים בפרספקטיבות חישוביות, ומשתדל לפשט היבטים סבוכים.

1.7. מדריך זה אינו מתיימר לטפל בשלל האפשרויות הגלומות במערכת, אלא רק בעיקריות שבהן. רוב הקוראים, יש

להניח, יוכלו להתחיל מן הנקודה המוצעת כאן ולהמשיך הלאה לבד בדרך של ניסוי ותעיה. המדריך הוא חלק קטן מנסיון כוללני מתמשך להנחת יסודות למחקר חישובי נגיש בספרות עברית, שפירותיו עתידים לראות אור בשנת 2018. ניסיון זה, ובתוכו גם כתיבת המדריך הנוכחי, לא היה מתאפשר בלי תמיכת מכון 'הקשרים' לחקר הספרות והתרבות היהודית והישראלית באוניברסיטת בן-גוריון בנגב. תודתי שלוחה לצוות המכון ולפרופ' יגאל שוורץ העומד בראשו, על הסיוע ועל האמון. וכמובן, תודה מעומק הלב לצוות CATMA, ולפרופ' יאן כריסטוף מייסטר העומד בראשו, על השיעור הגדול שלמדוני, ושעודני למד.

אייתי מרינברג-מיליקובסקי

המבורג, סיון תשע"ז / יוני 2017

2.1. כדי להפיק את מירב התועלת מן המדריך שלהלן, מומלץ לעקוב אחר השלבים הבאים תוך כדי התנסות ממשית בעבודה עם חומר גלם טקסטואלי אמיתי. לשם כך, יש להכין מראש קורפוס קטן – שני טקסטים קצרים מאוד (מפסקה בודדת ועד עמוד או שניים, אך לא יותר מכך). 'הכנה', בהקשר הנוכחי, פירושה התקנת עותק דיגיטלי של שני הטקסטים, שמורים בשני קבצי word נפרדים.¹

2.2. קבצי הטקסט צריכים להיות 'נקיים' לגמרי: גוף הטקסט בלבד, ללא כותרות והסברים. סימני פיסוק יכולים להישאר (CATMA מונה כל סימן כתו עצמאי, במנותק מן המילה הסמוכה לו). סימני ניקוד יכולים להישאר גם הם.² על מנת לעמוד על מלוא הפוטנציאל של המערכת, כדאי מאוד לבחור טקסטים מוכרים היטב למשתמש במסגרת עבודתו המחקרית הרגילה.

2.3. במסגרת המדריך יודגמו שלבי העבודה במערכת בניתוח ספרותי-נראטולוגי ראשוני וחלקי של שני סיפורים תלמודיים קצרים, שאחד מהם לפחות מוכר למדי אף מחוץ לגבולות הצרים של קהיליית חוקרי ספרות חז"ל. הבחירה בסיפורים תלמודיים, ובפרספקטיבה נראטולוגית, מקורה אך ורק בתחומי ההכשרה של כותב שורות אלה;³ הקורא מוזמן, כאמור, להתנסות ב-CATMA עם הטקסטים המעסיקים אותו, תהא אשר תהא שיטת הניתוח, יהיו אשר יהיו יעדיה.

¹ CATMA יודעת אמנם לייבא טקסטים גם ממקורות אחרים, ובתצורות דיגיטליות מגוונות; קוראים בעלי רקע מוזמנים להתנסות בכך. עם זאת, הנחת היסוד שלי היא שלמרבית החוקרים נוח במיוחד להשתמש דווקא בקבצי Word המאפשרים עיבוד מקדים בסביבה מוכרת. כך או כך, לא כאן המקום להציג את הדרכים השונות ליצירת טקסט דיגיטלי – אם בדרך של העתקה ידנית, סריקה מתוחכמת הכוללת זיהוי תווים (OCR), או השגת טקסט שכבר עבר דיגיטציה וזמין (באופן חוקי) למשתמש ברשת האינטרנט. היבטים חוקיים הנוגעים להגנה על זכויות יוצרים בטקסטים דיגיטליים אף הם לא יידונו כאן.

² במקרים נדירים יתכן שיבוש בקריאת מילים מנוקדות במערכת. הבעיה אמורה להתגלות בשלב מוקדם של תהליך העבודה. במצב כזה מומלץ להחליף את הקובץ המוזן למערכת בקובץ שהניקוד הוסר ממנו (כלים להסרת ניקוד מצויים באתרים שונים במרשתת).

³ דנתי בסיפורים אלה ובדומים להם בהקשר אחר ובכלים אחרים במקום אחר: איתי מרינברג-מיליקובסקי, לא ידענו מה היה לו: ספרות ומשמעות באגדה התלמודית, רמת-גן 2016.

(1) משה בבית מדרשו של רבי עקיבא (בבלי מנחות כט ע"ב)

אמר רב יהודה אמר רב: בשעה שעלה משה למרום מצאו להקדוש ברוך הוא שיושב וקושר כתרים לאותיות. אמר לפניו: רבונו של עולם, מי מעכב על ידך? אמר לו אדם אחד יש שעתיד להיות בסוף כמה דורות ועקיבא בן יוסף שמו, שעתיד לדרוש על כל קוץ וקוץ תילין תילין של הלכות. אמר לפניו: רבונו של עולם, הראהו לי! אמר לו: חזור לאחורך. הלך וישב בסוף שמונה שורות ולא היה יודע מה הן אומרים. תשש כחו. כיון שהגיע לדבר אחד אמרו לו תלמידיו: רבי, מנין לך? אמר להן: הלכה למשה מסיני. נתיישבה דעתו. חזר ובא לפני הקדוש ברוך הוא, אמר לפניו: רבונו של עולם, יש לך אדם כזה ואתה נותן תורה על ידי? אמר לו: שתוק, כך עלה במחשבה לפני. אמר לפניו: רבונו של עולם, הראיתי תורתו, הראני שכרו! אמר לו: חזור לאחורך. חזר לאחוריו. ראה ששוקלין בשרו במקולין. אמר לפניו: רבונו של עולם, זו תורה וזו שכרה? אמר לו: שתוק, כך עלה במחשבה לפני.

(2) משה לא אומר שלום לקב"ה (בבלי שבת, פט ע"א)

ואמר רבי יהושע בן לוי: בשעה שעלה משה למרום מצאו להקדוש ברוך הוא שהיה קושר כתרים לאותיות. אמר לו: משה, אין שלום בעירך? אמר לפניו: כלום יש עבד שנותן שלום לרבו? אמר לו: היה לך לעזרני. מיד אמר לו: 'ועתה יגדל נא כח ה' כאשר דברת'.

2.5. במקרה זה, שני הסיפורים הועתקו מפרוייקט 'ויקיטקסט' העברי – ספריה מקוונת חופשית של טקסטים שמספרם הולך וגדל בהתמדה.⁴ סימני פיסוק נוספו כדי להקל על הקורא, בשעה שקיצורים וראשי תיבות נפתחו (כך למשל 'ר' הפך ל'רבי'; רבש"ע הפך ל'רבונו של עולם'), ומראי מקום הוסרו (כך הוסרה הוסרה ההבהרה 'במדבר יד' שציינה את מקור הפסוק המצוטט בפי משה בסוף הסיפור השני). שמות הסיפורים הוצעו על ידי, הן על מנת לפשט את הדיון, הן על מנת להעניקם כשמות לשני הקבצים הנפרדים שבהם נשמרו הטקסטים.

2.6. לאחר התקנת הקורפוס אנו כמעט כשירים לגשת אל המערכת; אך מוטב שנתעכב רגע על ניסוח שאלת המחקר.

⁴ <https://he.wikisource.org/wiki>. נוסח התלמוד בפרוייקט ויקיטקסט תואם את נוסח הדפוס, ולצורך ענייננו כאן די בכך.

3.1. כפי שכבר הובהר, המערכת המוצגת כאן אינה מתרחקת יתר על המידה מן הפרקטיקות המסורתיות של מדעי הרוח. בניגוד לכלים אחרים שמציעים להתחיל את תהליך המחקר מ'משחק' חופשי בדפורמציות דיגיטליות של הטקסט – למשל, באמצעות חיתוכו ופריסתו מחדש בהתאם לתפוצת המילים המרכיבות אותו⁵ – CATMA מזמינה אותנו לחשוב תחילה מה אנחנו מעוניינים לבדוק. מאלינו מובן כי התהליך הסופי עתיד להיות מעגלי: דפורמציות דיגיטליות של הטקסט עשויות באמת לעורר שאלות חדשות לגמרי, כאלה שלא יכולנו לצפות להן מראש; אך לא כדאי לגשת אל הטקסט בלי איזושהי מחשבה ראשונית, אם לא שאלת מחקר של ממש.

3.2. ואולם, כיצד 'מתרגמים' את שאלת המחקר הרגילה – כולה, או חלק ממנה – לשפה חישובית/כמותית? הואיל ופעולה זו אינה שכיחה במיוחד במחקר השגרתי במדעי הרוח, מן הראוי לתת עליה את הדעת. אמנם, כפי שכותב פרנקו מורטי, "טקסטים עשויים ממילים, שורות, עמודים, ואפשר בהחלט למדוד אותם"⁶, אך איזה אלמנט מדיד יהיה בעל ערך מחקרי עבורנו? מה תהא תרומתה של המדידה, כאשר נבקש לשוב מן המספרים אל המשמעות?

3.3. את פעולת התרגום המתוארת לעיל נכנה, בעקבות מורטי, אופרציונליזציה (operationalization). מקור המושג במדעים המדויקים). המילה אמנם אינה מתגלגלת בנקל על הלשון, וחלופה עברית סבירה עבורה אינה בנמצא, אך היא מבטאת היטב את העקרון המתודולוגי המופשט העומד ביסוד כל מחקר חישובי. היא מתארת – מסביר מורטי בהשראת הפיזיקאי פרסי ברידג'מן –

את התהליך שבו קונספציות הופכות לסדרה של פעולות (operations), שמאפשרות, בתורן, למדוד אובייקטים מכל הסוגים. אופרציונליזציה פירושה בניית גשר מן הקונספט אל המדידה, ומשם אל העולם.

⁵ ראו למשל מערכת Voyant, וכן תיאורה במסגרת פרדיגמטית רחבה יותר במסגרת פרוייקט Hermeneuti.ca.

⁶ Franco Morretti, 'Operationalizing: Or, the Function of Measurement in Modern Literary Theory', Stanford

Literary Lab Pamphlet no.6 (December 2013), p.2. (א.מ.) וכן להלן בכל מקום אחר במדריך.

במקרה שלנו: מקונספציות של תיאוריה ספרותית, דרך כמה צורות של כימות, ומשם אל הטקסטים

הספרותיים.⁷

3.4. במילים פשוטות, איננו מודדים דברים סתם כך. פעולת המדידה אמורה לשרת תכלית קונספטואלית מסוימת; וכאשר נותנים למספרים לדבר, היא נעשית כמעט זהה עמה. שכן, כפי שמדגים ברידג'מן (בקטע המצוטט אצל מורטי, שם), "כדי לגלות את אורכו של אובייקט עלינו לערוך כמה פעולות פיזיקליות. הקונספט 'אורך' מתמלא, אפוא, כאשר מתמלאות האופרציות שבעזרתן נקבע אורכו של דבר. כלומר, הקונספט 'אורך' כרוך בסדרה של אופרציות שבאמצעותן אורך נקבע, ולא בדבר זולתן". כך, אם שאלת המחקר המסורתית שלנו היא הקונספציה שאותה אנו מעוניינים לבדוק, או הרעיון שאת גילומיו הכמותיים אנו עומדים לחפש, הרי 'אופרציונליזציה' שלה תהיה הנסיון להבהיר ולנסח לעצמנו את הקשר בינה ובין המדידה הצפויה.

3.5. בתהליך המחקר החישובי נגלה לא אחת כי אתגר ההבהרה והניסוח עומד לפתחנו גם בכיוון ההפוך: מדדנו דבר-מה, הרגשנו שמדידה זו רלוונטית עבורנו מסיבה כלשהי, אך אנו מתקשים עדיין לעמוד על מלוא משמעותה. מכל מקום, כפי שמבהירה בצדק אוולין גיוס, "אופרציונליזציה איננה משימה קלה, שכן קונספציות תיאורטיות רבות, כמו גם המשגות רבות, נותרות מעורפלות או מופשטות מכדי לאפשר פורמליזציה ישירה".⁸

3.6. מערכת CATMA, נראה בהמשך, כוללת כמה אפשרויות שעשויות לפשט במקצת את האתגר. בשונה ממערכות רבות אחרות המזוהות עם מדעי הרוח הדיגיטליים, היא מודדת לא רק את ה-data, כלומר, את החומר הגולמי, את הנתונים הטקסטואליים האובייקטיביים (מילים, צירופי מילים וכיוצא בזה), שבאמת אינם נענים תמיד בנקל לשאלת מחקר מופשטת, אלא גם את ה-metadata, כלומר, את המשמעויות שהחוקר מייחס לטקסט באופן סובייקטיבי, ושאותן הוא מתעד על גבי הטקסט באמצעות תגים מגוונים. כך, שאלת המשמעות של האובייקט הנמדד – או העתיד להימדד – מלווה את המשתמש מן השלבים המוקדמים ביותר של העבודה, כאשר הוא נדרש להגדיר ולעצב את ספריית התגים הרלוונטית עבורו. במהלך העבודה, בכל רגע נתון, הוא יכול לחזור ולשנות את הגדרות מערכת התגים, ואפילו לתעד על פני הטקסט את ההתלבטות המלווה אותו בקריאת הטקסט וניתוחו.

⁷ מורטי, שם, עמ' 1 [ההדגשה שלי – א.מ.], וכן להלן.

⁸ Evelyn Gius, ['Narration and Escalation: An Empirical Study of Conflict Narratives'](#), *Diegesis* 5.1 (2016), p. 10.

3.7. על מנת שלא להותיר את הדיון במישור התיאורטי בלבד, הבה ניטול לדוגמא את הקורפוס שהוצע קודם. לפנינו שני סיפורים תלמודיים. לשניהם גיבור משותף (משה רבנו), סיטואציה דומה (משה עולה למרום כדי לקבל שם את התורה), ואפילו מטבעות לשון זהות ("בשעה שעלה משה למרום", "יושב וקושר כתרים לאותיות"). אך אם שאלת המחקר שלנו – אותה נגדיר בינתיים באופן רופף למדי – היא, נניח, 'מהן אבני הבניין של הסיפורים', או 'מהי הליבה העלילתית של הסיפורים, ומה מקומה בין שלל האירועים המתרחשים בהם', לא נוכל להסתפק רק במניית המילים המשותפות; אנו נתקשה מאוד למצוא קשר ישיר או עקיף בין חומר הגלם המילולי, הדל מצד עצמו, למה שבאמת מעסיק אותנו. לעומת זאת, אם נבנה סט תגים לסיווג הדמויות וסט תגים נוסף לסיווג האירועים בעלילה, ונתייג בעזרתם את הקורפוס, נוכל לשאול את המחשב לא רק 'כמה פעמים (והיכן בדיוק) מופיע השם 'משה' בקורפוס?' – שאלה שהמחשב יודע לענות עליה בנקל, גם אם אינו יודע דבר על משה זה – אלא גם, נאמר, 'כמה פעמים (והיכן בדיוק) מתרחשים בסיפורים אירועים בעלי חשיבות מבחינה עלילתית'. המחשב, כמובן, אינו יודע מהו 'אירוע' ומה עושה אותו ל'בעל חשיבות מבחינה עלילתית'; ובניגוד ל'משה', אין למילים אלה זכר בטקסטים. אך לאחר שנתייג את הטקסט בהתאם לפרמטרים הנגזרים מן האופרציונליזציה של שאלת המחקר, יוכל המחשב למנות ולמפות את התגים שהצענו, ולספק – יש לקוות – תשובות מעניינות.

3.8. אם נרצה לשנות תוך כדי מחקר את הגדרות היסוד שלנו ל'אירוע' ו'עלילה', או לחילופין אם נרצה לשקול בשלב כלשהו תובנות העולות מתפיסה אחרת לגמרי של הטקסט – שלנו או של אחרים – נוכל לעשות זאת בנקל, בלי לוותר בהכרח על המחשבות הראשונות, ובלי למנוע אפשרות לעדכון נוסף בעתיד. מילות המפתח כאן הן 'רב-שכבתיות' מצד אחד, ו'**טקסטואליות**' מצד שני: כל מידע, כל מבט, יכולים להיות מתועדים בשכבה נפרדת על גבי הטקסט, כל עוד אפשר לגזור קשר ישיר או עקיף ביניהם לבין הטקסט עצמו. אם נרצה, נוכל להתעלם משכבות שונות בשלבם שונים של הניתוח, או לחילופין להקים לתחייה רעיונות ישנים שפסלנו בעבר. כל אלה תהליכים מוכרים לכל מי שעוסק במחקר; הם רק מקבלים כאן קונפיגורציה חדשה.

3.9. שתי הערות ביקורתיות – שתיים לפחות – עשויות לעלות כעת. ראשית, מדוע עלינו לטרוח בלימוד תוכנה חדשה ומורכבת, המציעה לנו בסך הכל גרסה מסובכת לפעולה אנושית פשוטה, שאפשר לעשותה באותה מידה בעט סימון זוהר על גבי דף נייר? שנית, מה באשר לטעויותיו של החוקר, או למצער להכרעותיו השנויות-במחלוקת, המתועדות גם הן באותה דרך? האם איננו מגייסים את המחשב לשירותנו כדי להגיע, לכאורה, לתשובות 'נכונות' יותר, 'מדעיות' יותר, 'אובייקטיביות' יותר, או לפחות כאלה המעוגנות בממצאים אמפיריים מובהקים? לשתי הערות אלה תשובה עקרונית אחת. CATMA, אכן, אינה מעודדת תפיסה נאיבית של ה'גולם' המודרני, ואינה מייחסת לו יכולות בלתי-מוגבלות – ודאי בכל הנוגע ליכולות הרמנויטיות; ואולם, ככל שמערכת התגים מסועפת

יותר, רב-רובדית ועשירה יותר – או, לחילופין, ככל שהיקף הטקסט או הקורפוס הנבדק גדול יותר – כך מתגלה יתרוננו של המחשב על האדם. בשלב כלשהו כמות הנתונים שיש לעבד ולצרף יחד עלולה לעלות על כושר ההשגה שלנו, בעוד עבור המחשב אין היא אלא סדרה של מספרים בזכרון אינסופי. בה בעת, המחשב – לפחות בסביבה המוצגת כאן – אינו מייחס לעצמו יוקרה פרשנית עצמאית; הוא אינו מתכחש להישגי הקריאה של החוקר, אלא מסייע לו להעמיד אותם לבחינה יסודית. מערכת CATMA מעודדת את המשתמש לעיון מדוקדק בטקסט, משום שעליו להראות שוב ושוב היכן וכיצד באות לידי ביטוי התופעות המעסיקות אותו. תיאוריה נזילה מדי – ספק אם תצלח כאן; והואיל והטקסט לא תמיד נענה בנקל לציפיות המוקדמות שתלו בו, נמצא החוקר מתקן ומעדכן את הנחות היסוד עימן יצא לדרך. במובן זה, ולא רק בו, מציעה המערכת המתוארת כאן חידוש לא מבוטל.

3.10. הקורא מוזמן אפוא לנסח כעת את שאלת המחקר שאותה הוא מבקש לבחון בראי הקורפוס שהתקין, ולנסות 'לתרגם' אותה לשפה חישובית/כמותית. לשם כך, ובהתאמה לאפשרויות שמערכת CATMA עתידה להציע בפניו, עליו לחשוב בעיקר בשני כיוונים:

3.10.1. ראשית, עליו לחשוב על האלמנטים הטקסטואליים ה'אובייקטיביים', אלו הנמצאים בשכבת ה-**data** של הטקסט או הקורפוס, שיש להם חשיבות למחקרו;

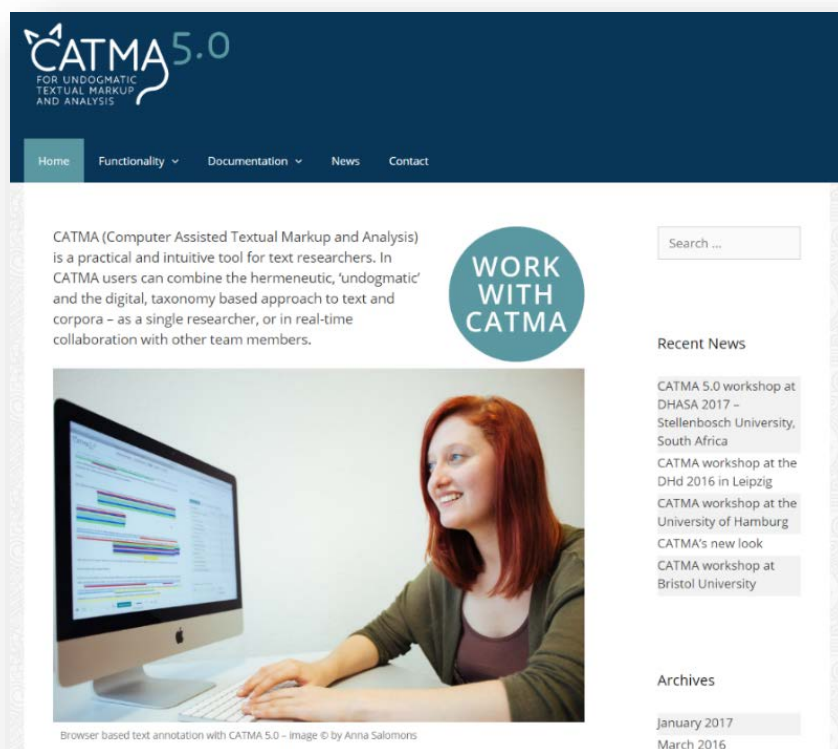
3.10.2. שנית, עליו לחשוב על התופעות, או הקטגוריות הפרשניות, שעימן ייגש אל הטקסט או הקורפוס, ושאותן יבקש לתייג ולסמן על גביו בשכבת ה-**metadata**.

3.11. מכאן ואילך ינוע המדריך לאורך ציר שתחילתו בטיפול ב-**data** גולמי, המשכו ביצירת **metadata**, וסופו בניתוח משולב של **data** ו-**metadata** יחד.



4.1. על מנת להיכנס למערכת נגלוש תחילה לכתובת catma.de.

4.2. כך נראה דף הבית של המערכת. עמוד זה הוא השער החיצוני למערכת: הוא מכיל מידע כללי על אודותיה, חומרי הדרכה ואימון, חדשות ועדכונים, וכל כיוצא בזה. וחשוב מכל – הוא מכיל את כפתור הכניסה למערכת.



4.3. בשלב ראשון, המשתמש המתחיל עשוי למצוא תועלת בעיקר במידע שתחת הלשונית Functionality, שם מתוארים רכיבי המערכת וההגיון המבני והקונספטואלי של כל אחד מהם. קוראי מדריך זה מוזמנים להמשיך ישר לעבודה במערכת, כשהם מלווים בהנחיות שלהלן; אך מומלץ מאוד לחזור ולעיין מדי פעם גם בחומרי ההדרכה שבאתר.

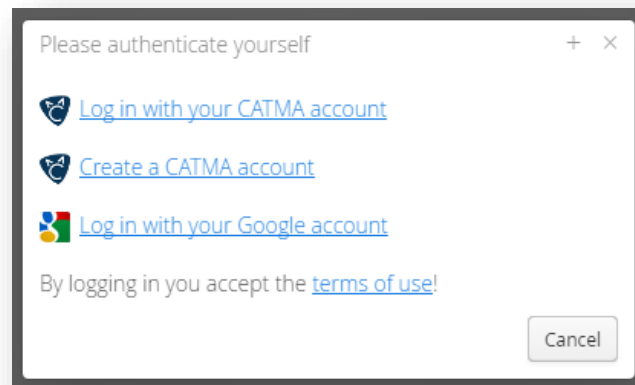


4.4. כדי להתחיל לעבוד, נלחץ על כפתור WORK WITH CATMA הנתון במסגרת

עגולה וממוקם סמוך לחלקו העליון של המסך.

4.5. בשלב זה נגיע למסך לבן חלק כמעט לגמרי, שבמרכזו הזמנה באפור: please log in to get started.

4.6. לחיצה על כפתור sign in, הממוקם בפינה הימנית עליונה של המסך, תוביל אותנו אל החלון הבא:



4.7. בכניסה ראשונה למערכת יהא עלינו להירשם – אם באמצעות החשבון שלנו בגוגל (Log in with your

Google account), ואם, לו רצוננו להרחיק את גוגל ממחקרנו, באמצעות רישום ישיר באתר המערכת

(Create a CATMA account). הרישום הראשוני אורך דקה או שתיים.

4.8. מכאן ואילך נוכל להיכנס תמיד אל המערכת או באמצעות חשבוננו בגוגל, או באמצעות כניסה לחשבון

CATMA שיצרנו (Log in with your CATMA account). לאחר שנתייג טקסט נמצא כי לכניסה הבאה

אל המערכת כמה מאפיינים ייחודיים, המאפשרים להמשיך לעבוד מן הנקודה בה הפסקנו. מאפיינים אלה יתוארו

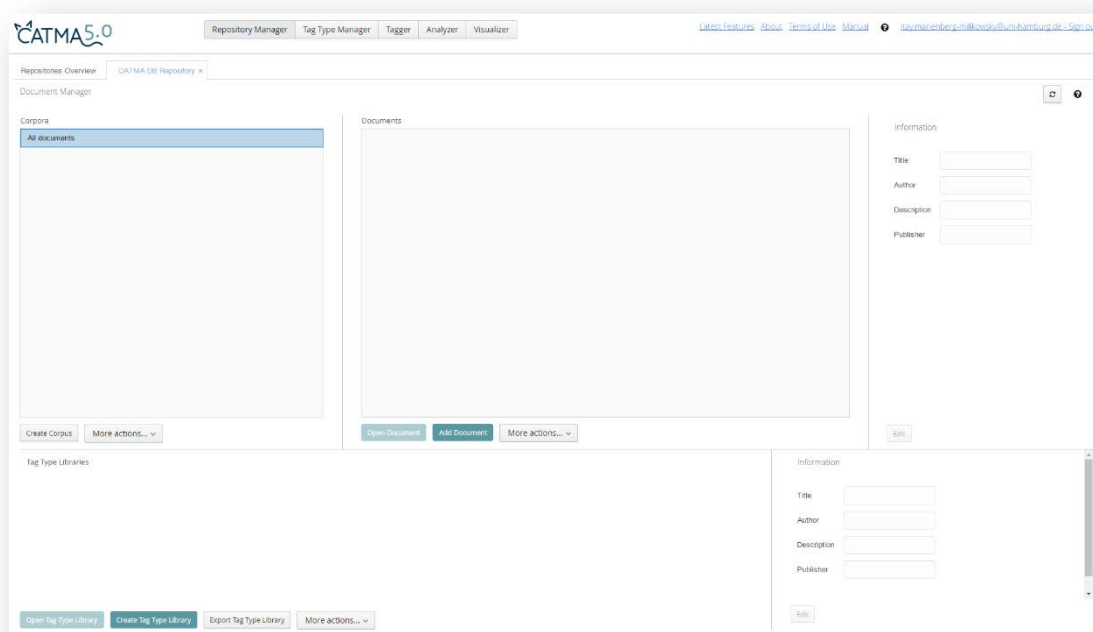
[בהמשך](#).

4.9. כפי שנראה מיד, המערכת מציעה בפנינו סביבת עבודה מלאה לטיפול ממושך ומקיף בטקסטים, ולא רק כלי

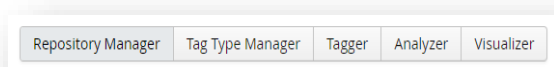
שנדרשים אליו מעת לעת על מנת לפתור בעיה מקומית. כל עבודה שנעשה במערכת עתידה להשמר בחשבוננו

בשרתי המערכת שבאוניברסיטת המבורג, וכמו כן להיות מגובה בשלושה מקומות נוספים.

5.1. כעת אנו במסך הכניסה ה'אמיתי' של המערכת – או ליתר דיוק, במסך הכניסה של החשבון הפרטי שלנו בתוכה. בשלב זה רובו לבן, אפור וריק, שכן עדיין לא הזנו מידע לתוכו. ככל שנעבוד יותר ב-CATMA, כך מסך זה יצבור לתוכו חומרים נוספים.



5.2. נתבונן בחמש הלשוניות בחלקו העליון:



Repository Manager (א)

Tag Type Manager (ב)

Tagger (ג)

Analyzer (ד)

Visualizer (ה)

לשוניות אלה מאפשרות לנו לעבור בין חמישה מסכים שונים, המשקפים את חמש סכיבות העבודה המוצעות באופן סינכרוני במערכת. בשלב הנוכחי, בטרם נתחיל לעבוד, המסכים עדיין אינם מציגים את הפונקציות המגוונות הזמינות בכל אחד מהם. בהמשך נלמד כיצד לעשותם פעילים.

5.3. מטרת הסקירה להלן היא להציע מבט ראשוני בלבד על רכיבי המערכת (מבט מעמיק יותר יוצע בפרקים הבאים):

5.3.1 **Repository Manager** – נקודת המוצא לעבודה במערכת. במרכז המסך אנו עתידים לראות את כלל

הטקסטים שנעלה לתוכנה; בצדו השמאלי נמצא חלוקה של הטקסטים לקורפוסים; ובחלקו התחתון נמצא את רשימת ספריות התגים שנגבש בהמשך. במסך זה נזין טקסטים למערכת, 'נטען' את הטקסט או הקורפוס שבו ברצוננו לעסוק, ואף נגדיר את הכותרות הראשיות של ספריות התגים.

5.3.2 **Tag Type Manager** – במסך זה ננהל את ספריות התגים שלנו, ללא כל תלות בטקסט מסוים. כאן

נגדיר אילו תופעות ברצוננו לסמן, כאן נסווג אותן, כאן נארגן אותן במבנה אופקי או אנכי (הירארכי), כאן נקבע את הצבעים שיצינו אותן – אך לא נתייג בפועל. במילים אחרות, כאן המקום לחשוב באופן מופשט על שאלת המחקר ועל היבטים מרכזיים באופרציונליזציה החשובית שלה.

5.3.3 **Tagger** – במסך זה יופיע הטקסט שאותו אנו מבקשים לחקור, ושאותו העלנו למערכת. בצד הטקסט

תופיע מערכת תגים לפי בחירתנו, וכמו כן יצויין שמה של שכבת ה- metadata שבתוכה יישמרו תיוגינו. במסך זה נקרא את הטקסט ברצף, וכאן נתייג אותו הלכה למעשה. במסך זה נבלה כנראה את משך הזמן הארוך ביותר בעבודתנו במערכת.

5.3.4 **Analyzer** – במסך זה נקבל פריסה של הטקסט או של הקורפוס בהתאם לפרמטרים הנתונים לבחירתנו

– מאוצר המילים הבסיסי של הטקסט או הקורפוס, ועד פילוח מתקדם של תופעות מגוונות ב- data או ב- metadata. כאן נוכל להגדיר שאילתות ברמת מורכבות משתנה, ולעיין מחדש במקטעים שונים של הטקסט בהתאם לבחירתנו. במידת הצורך, נוכל להוסיף תיוג לחלקים מן הטקסט ישירות ממסך זה, ואפילו לתייג קטעי טקסט שונים ורחוקים זה מזה בבת אחת. כאן נוכל גם לבקש מהמערכת להפיק עבורנו גרפים מסוגים שונים.

5.3.5 **Visualizer** – במסך זה יוצגו הגרפים שנבקש במסך הקודם. כאן נוכל לבחון אותם, ובמידת הצורך

'לשחק' מעט בהגדרות של כל אחד מהם על מנת להתאימו לצרכינו המחקריים.

5.4. יושם אל לב, כי המערכת אינה כופה על המשתמש סדר עבודה ליניארי מסוים (למעט כמה שלבים טכניים שיתוארו בהמשך). בשלב העיקרי של המחקר, המשתמש יכול לעבור בחופשיות רבה בין המסכים השונים על מנת לשלב את הפונקציות המוצעות בהם בתהליך מחקרי אינטגרטיבי שבו פרודורות שונות משפיעות זו על זו, מתקנות זו את זו, ואף מאתגרות זו את זו.

5.5. בהמשך נלמד כי ההפרדה בין המסכים השונים אינה חדה לגמרי: כך למשל יכול המשתמש להשפיע על עיצוב ספריית התגים לא רק מתוך המסך המוקדש לכך, אלא גם מתוך המסך שבו הוא מתייג את הטקסט בפועל; כך למשל יכול המשתמש לתייג בבת אחת מקטעי טקסט רבים שקובצו יחד באופן אוטומטי במסך הניתוח.

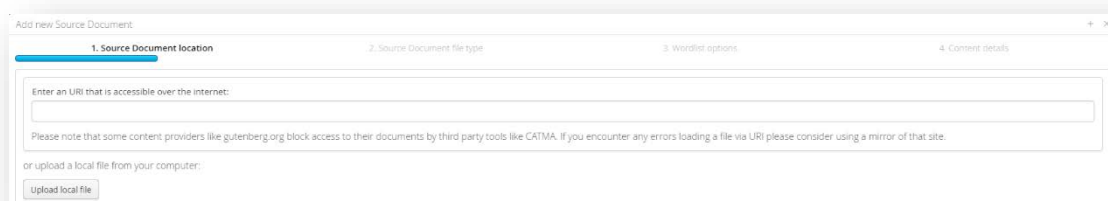


5.6. כמעט בכל מסך ימצא המשתמש צלמית קטנה (אחת או יותר) המוצגת כסימן שאלה. לחיצה על הצלמית תפתח חלון עזרה ובו הסבר טכני חלקי וקצר על ייעודו של האזור הנבחר במסך, על הפונקציות השונות המתאפשרות בו, ועל הדרכים להפעלתן. כדאי מאוד להיעזר בו במקרה הצורך.

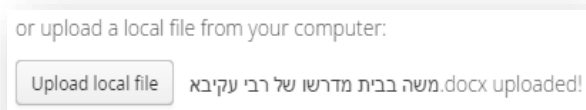
5.7. בסיום העבודה יש לצאת מן המערכת בלחיצה על כפתור ה- sign out שבפינה הימנית-עליונה. בדרך כלל המערכת שומרת את המידע המתווסף לחשבון באופן אוטומטי, אך ליתר בטחון מומלץ להקפיד על יציאה מסודרת.

5.8. בפרקים הבאים נתאר תהליך עבודה שלם בזעיר אנפין: מהזנת הטקסט וארגון הקורפוס, דרך ניתוח אוטומטי של חומר גולמי, התקנת מערכת תגים ושימוש בה, ועד עיבוד משולב של ה'חשובי' וה'אנושי', ואף הפקת גרפים. כך נתוודע לפונקציות העיקריות בכל מסך. הקורא מוזמן לעקוב אחר ההדגמה (המתבססת, כאמור, על שני הסיפורים התלמודיים שהוצגו לעיל), וללוות אותה צעד-צעד בפרוייקט משלו, בהתאם לטקסטים שבחר, ולשאלת המחקר המעסיקה אותו.

- 6.1. כאשר קבצי הטקסטים מוכנים, אפשר להזין אותם למערכת. לשם כך נלחץ על כפתור Add Document, הנמצא בערך במרכז מסך ה-Repository Manager. נקבל את החלון הבא:



- 6.2. CATMA מאפשרת להזין טקסט ישירות מן המרשתת, אך אנו נבחר באפשרות השנייה – העלאת קובץ מקומי מן המחשב (Upload local file). זאת, משום שעבור חוקרים רבים מדובר בדרך פורייה ובטוחה יותר: היא מאפשרת התקנה ועיבוד מוקדמים של הטקסט בהתאם לצורך ולעניין, לפני שעורכים עליו מניפולציות נוספות.



- 6.3. לאחר בחירת הקובץ וטעינתו למערכת, נקבל חיווי על כך:



- להשלמת התהליך נלחץ על כפתור NEXT בפינה הימנית-תחתונה של המסך.

- 6.4. במסך הבא נוכל לוודא שאכן הטקסט שהעלינו התקבל בסוג הקובץ המתאים. הואיל ו-CATMA מזהה היטב קבצי word, כמעט תמיד נוכל להמשיך לשלב הבא באמצעות לחיצה נוספת על כפתור NEXT.

6.5. כעת עלינו לסייע

למערכת לזהות

את שפת הטקסט.

המחשב, משום

מה, אולי זיהה

את שפתנו

כיפנית (!) – jp-Japan – אך אפשר לעזור לו בנקל. אם נכתוב

Heb במקום המתאים, נקבל שתי אפשרויות: Hebräisch ו-

Hebräisch-Israel. לצורך ענייננו כאן אין הבדל של ממש

בין שתי ההגדרות, ובחירת כל אחת מהן תהיה מספקת בהחלט.

בשלב הנוכחי נתעלם גם מן האפשרות לטיפול מתקדם יותר

ברשימת המילים, המוצע בצדו הימני של המסך, ובמקומו נלחץ שוב על NEXT.

6.6. בחלון הבא נשלים פרטים אחרונים ברישומי המסודר של הטקסט במערכת (מחבר, תיאור, מוציא לאור). נקפיד

לציין פרטים שיקלו עלינו לנהל בהמשך את מאגר הטקסטים שלנו – למשל, כך:

6.7. בגמר הרישום, נלחץ על כפתור Finish בפינה הימנית-תחתונה של המסך – וכעת הטקסט יופיע ברשימת

המסמכים במסך הראשי.

6.8. בדרך כלל נמצא תועלת בשיוך טקסט לקורפוס –

פעולה שתאפשר לנו, בהמשך, לנתח את כל הטקסטים

בקורפוס מוגדר בבת אחת. בשלב הנוכחי, הטקסט

שהכנסנו למערכת נמצא בקורפוס הכללי ששמו 'All'

'documents', ושאליו עתיד להגיע כל טקסט שנכניס אי פעם למערכת.

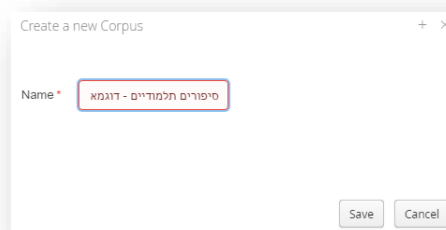
6.9. כדי להוסיף קורפוס ממוקד יותר, נפנה לחלקו השמאלי של מסך ה- Repository Manager, ומתחת לרשימת

ה- Corpora נלחץ על כפתור Create Corpus.

6.10. בחלון החדש שנפתח נגדיר את שם הקורפוס – למשל,

‘סיפורים תלמודיים – דוגמא לצורך התנסות’. לאחר לחיצה על

Save יופיע הקורפוס החדש בחלון ה- Corpora.

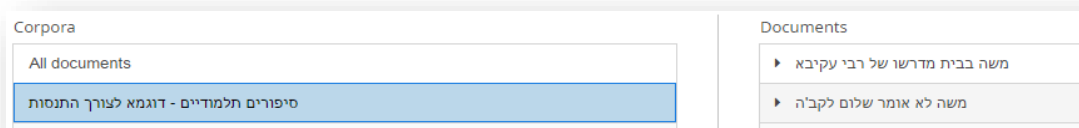


6.11. עתה עלינו להעביר את הטקסט שכבר הוזן למערכת אל

הקורפוס החדש. לשם כך, נלחץ על הטקסט המבוקש עם הסמן, ו'נגרור' אותו אל הקורפוס שיצרנו.

6.12. להשלמת הכנת התשתית של ההתנסות המוצעת כאן, נוסיף את הטקסט השני לקורפוס, בהתאם להנחיות דלעיל.

בסוף התהליך אמורים להופיע שני טקסטים במערכת, משויכים לקורפוס שהוגדר קודם.



6.13. יושם אל לב: המערכת המתוארת כאן מייחסת למושג 'קורפוס' משמעות אינסטרומנטלית גרידא, השונה במידה

רבה מן המשמעות המיוחסת לו בשיח הפילולוגי הרגיל. קורפוס, לדידה של CATMA, הוא **קבוצת טקסטים**

שברצוננו לחקור בבת אחת. המכנה המשותף היחיד שלהם הוא הצורך המחקרי שלנו. אם אנו עורכים מחקר

המוקדש, נניח, לעשרים טקסטים, נוכל לקבץ אותם יחד לקורפוס אחד; אך אם נבקש להתמקד בשלב כלשהו

בבחינה משווה של שניים מהם (או יותר), יהיה עלינו להגדיר קורפוס ייעודי לכך, ולגרור אליו את הטקסטים.

גרירה זו אינה גורעת את הטקסטים מן הקורפוס הקודם, ואינה משנה דבר בשכבת ה- metadata שנבנתה זה

מכבר על גבי כל אחד מהם; כל תכליתה היא לאפשר לנו לערוך פרוצדורות אוטומטיות מסוימות בפעולה מהירה

אחת, המטפלת בו-זמנית בקבוצת הטקסטים שהוגדרה על ידינו, ורק בה.

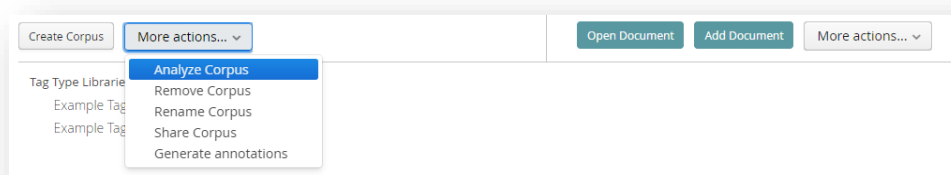
6.14. בפרקים הבאים נלמד כיצד להתחיל לנתח את הקורפוס בדרכים שונות: תחילה באופן אוטומטי, לאחר מכן באופן

ידני, ולבסוף בשילוב שתי הצורות.

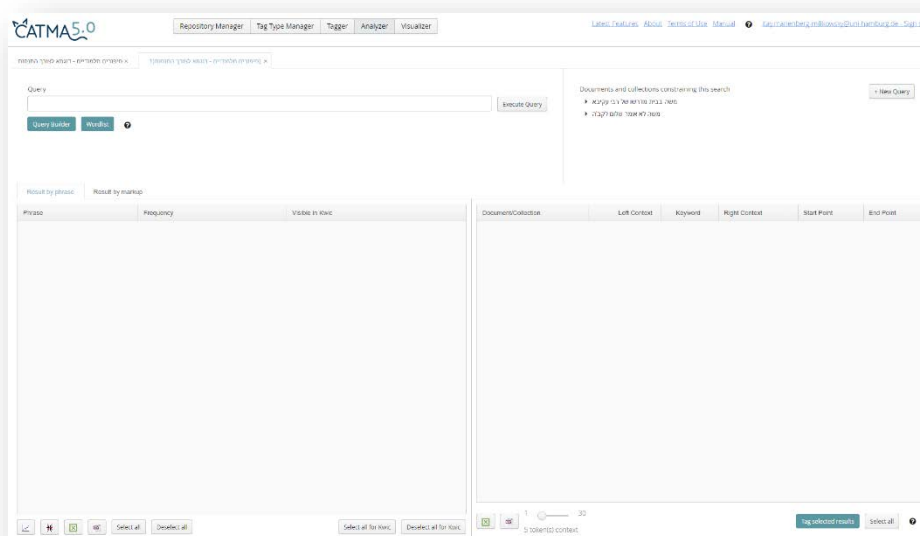
7. ניתוח אוטומטי ראשוני של טקסט או של קורפוס שלא תויגו

7.1. עיקר ייחודה של מערכת CATMA טמון, כפי שכבר נזכר לעיל, באפשרות לתיוג מושכל של הטקסט. אולם מחקרים רבים עשויים להיות נשכרים גם מניתוח אוטומטי של הטקסט או הקורפוס כשהם לעצמם, ללא כל שכבת תיוג. לשם כך, נסמן כעת את הטקסט או הקורפוס שברצוננו לבדוק.

7.2. במסך ה- Repository Manager, מתחת לחלונות ה- Corpora וה- Documents, נמצא כפתור More actions. לחיצה עליו תפתח בפנינו סדרת אפשרויות, ובראשן Analyze corpus, או Analyze document (בהתאמה).

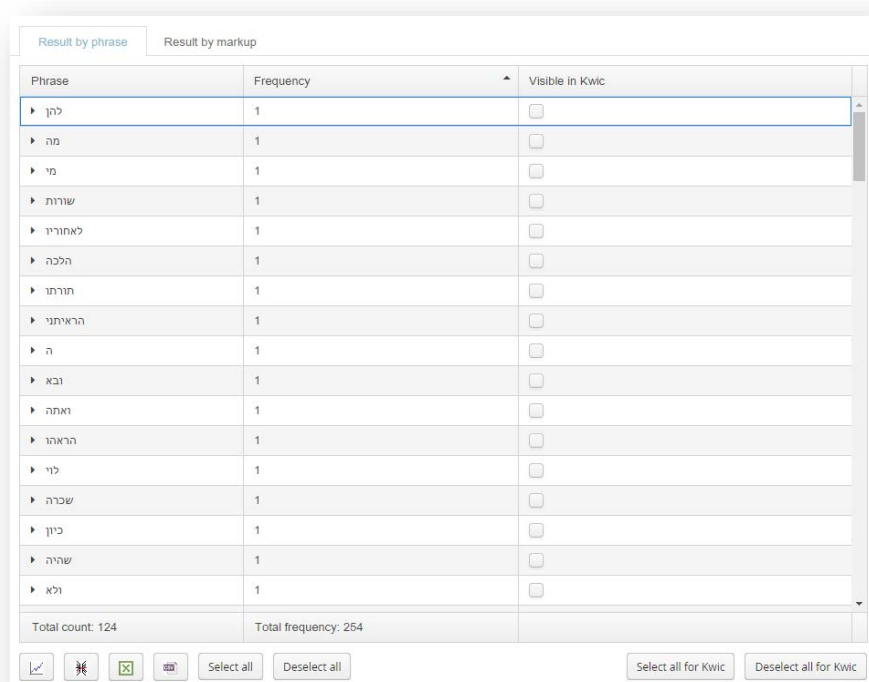


7.3. בחירה באפשרות זו תוביל אותנו לראשונה למסך ה- Analyzer. בצידו הימני-עליון של המסך נמצא את רשימת הטקסטים העומדים לניתוח; בצידו השמאלי-עליון של המסך נמצא את אזור בניית השאלות (Query). עד מהרה נכיר גם את שאר חלקי המסך.



7.4. הואיל ועדיין לא תייגנו את הטקסט, ניתוחנו חייב להתמקד לפי שעה ב- data בלבד, כלומר, בחומר הגלם הטקסטואלי כמות שהוא. פעולה ראשונה, מתבקשת מאליה, תהיה לבדוק את אוצר המילים. לשם כך נלחץ על כפתור Wordlist הנמצא מתחת לחלון ה- Qeury.

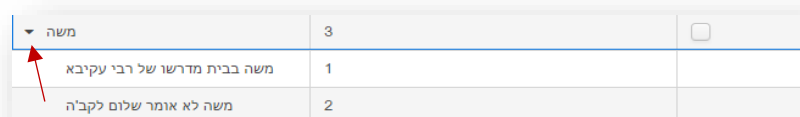
7.5. כעת נקבל את רשימת המילים המרכיבות את הקורפוס באזור השמאלי-תחתון של המסך. באזור זה מרוכזות תמיד כל תוצאות החיפוש, יהיו אשר יהיו יעדיו.



Phrase	Frequency	Visible in Kwic
להן	1	<input type="checkbox"/>
מה	1	<input type="checkbox"/>
מי	1	<input type="checkbox"/>
שורות	1	<input type="checkbox"/>
לאחוריו	1	<input type="checkbox"/>
הלכה	1	<input type="checkbox"/>
תורתו	1	<input type="checkbox"/>
הראיתי	1	<input type="checkbox"/>
ה	1	<input type="checkbox"/>
ובא	1	<input type="checkbox"/>
ואתה	1	<input type="checkbox"/>
הראהו	1	<input type="checkbox"/>
לוי	1	<input type="checkbox"/>
שכרה	1	<input type="checkbox"/>
כיון	1	<input type="checkbox"/>
שהיה	1	<input type="checkbox"/>
ולא	1	<input type="checkbox"/>
Total count: 124	Total frequency: 254	

7.6. לחיצה על כל אחת מכותרות העמודות תארגן את הטבלה מחדש: לחיצה על Phrase תערוך מחדש את התוצאות על פי סדר אלפביתי, מתווים שאינם אותיות (כגון סימני פיסוק), ועד האות ת' (לחיצה נוספת על אותו כפתור תארגן את התוצאות על פי סדר אלפביתי מהופך); לחיצה על Frequency תארגן את התוצאות על פי מידת התפוצה שלהן – מהנדירה ביותר אל הנפוצה ביותר (לחיצה נוספת על אותו כפתור תארגן אותם במהופך).

7.7. ליד כל מילה יש משולש קטן. לחיצה עליו מפרטת את היקרויות המילה בטקסטים השונים. כך למשל המילה 'משה' מופיעה שלוש פעמים בקורפוס – פעם אחת בסיפור שכותרתו 'משה בבית מדרשו של רבי עקיבא',



משה	3	<input type="checkbox"/>
משה בבית מדרשו של רבי עקיבא	1	
משה לא אומר שלום לקב"ה	2	

ופעמיים בסיפור שכותרתו 'משה לא אומר שלום לקב"ה':

7.8. סימון $\sqrt{\quad}$ בריבוע הקטן שמימין –

משה	3	<input type="checkbox"/>
משה בבית מדרשו של רבי עקיבא	1	
משה לא אומר שלום לקב"ה	2	

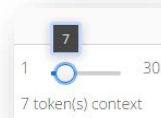
הנמצא בעמודה שכותרתה Visible

in Kwic, תציג את שלוש היקרויות

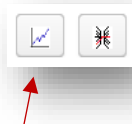
המילה הנבחרת – בהקשריה השונים – בצידו הימני-תחתון של המסך, בטבלת ה- Kwic (= Key Words In Context).

Document/Collection	Left Context	Keyword	Right Context	Start Point	End Point
משה בבית מדרשו של רבי עקיבא	למרום מצאו להקדוש ברוך הוא	משה	אמר רב: בשעה שעלה	31	34
משה לא אומר שלום לקב"ה	למרום מצאו להקדוש ברוך הוא	משה	בן לוי: בשעה שעלה	33	36
משה לא אומר שלום לקב"ה	אין שלום בעירך?	משה	לאותיות. אמר לו:	97	100

7.9. הצגה זו מאפשרת לעיין במילים המבוקשות בהקשרן, על מנת ללמוד כיצד הן מתפקדות בטקסט, מבלי לעיין בכולו. היקף ההקשר המובא לפני ואחרי המילה המבוקשת נתון לשינוי: ברירת המחדל של הטבלה היא להציג חמש מילים מכל כיוון, אך כפתור קטן בתחתית הטבלה מאפשר לשנות זאת ממילה אחת ועד שלושים מילים.



7.10. מסך ה- Analyzer מאפשר לנו גם להפיק שני סוגים של גרפים:



(א) גרף תפוצה (distribution graph) – מציג את תפוצתה של תופעה נבחרת, או

תופעות נבחרות, לאורך הטקסט. התופעה יכולה להיות ברמת ה- data ו/או ה-

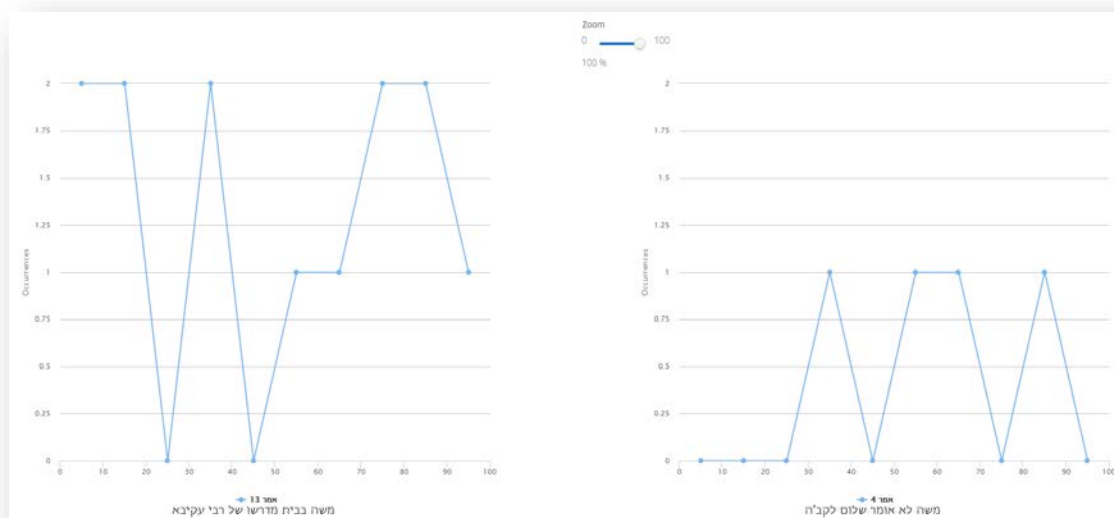
metadata.

(ב) גרף 'עץ כפול' (double tree) – מציג את התנהגותן ההקשרית של מילים נבחרות ברצף הטקסט.

7.11. כדי להפיקם ניגש אל צמד הכפתורים המיועדים לכך, וממוקמים תחת צידה השמאלי של רשימת הממצאים.

לחיצה על כל אחד מהם תיצור גרף שיופיע בחלון ה- Visualizer.

7.12. אם נסמן מילה ברשימת המילים, ונלחץ על הכפתור השמאלי, נקבל גרף תפוצה: במקרה זה הוא יציג את תפוצת המילה שבחרנו לאורך הטקסט או הקורפוס. כך למשל נראית תפוצת המילה 'אמר' בקורפוס המודגם כאן:

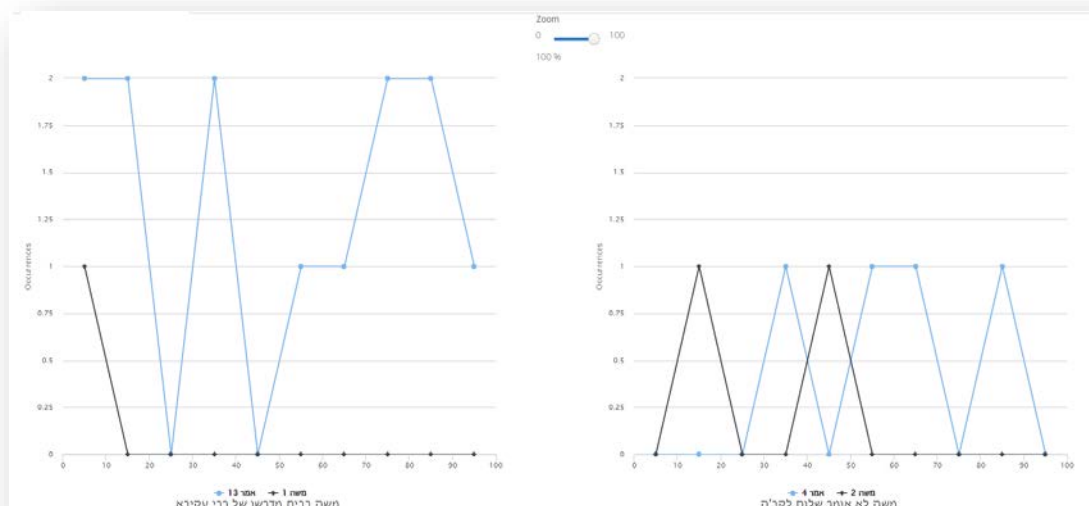


7.13. כך יש לקרוא את מסך היוזואליזציה: הואיל ובקורפוס שלנו יש שני טקסטים, המערכת הפיקה שני גרפים – אחד לכל טקסט (שם הטקסט מצוין בתחתית כל גרף, ומעליו המילה הנבחרת). לוי עבדנו עם טקסט אחד, היינו מקבלים גרף אחד; לוי עבדנו עם עשרים טקסטים, היינו מקבלים עשרים גרפים. הציר האופקי בכל גרף מחלק את הטקסט למאה חלקים שווים בעשרה מקטעים; הציר האנכי מציין את מספר ההיקריות של המילה (או התופעה) הנבחרת. עקומת הגרף משקפת אפוא את מספר ההיקריות של המילה בכל מקטע נתון.

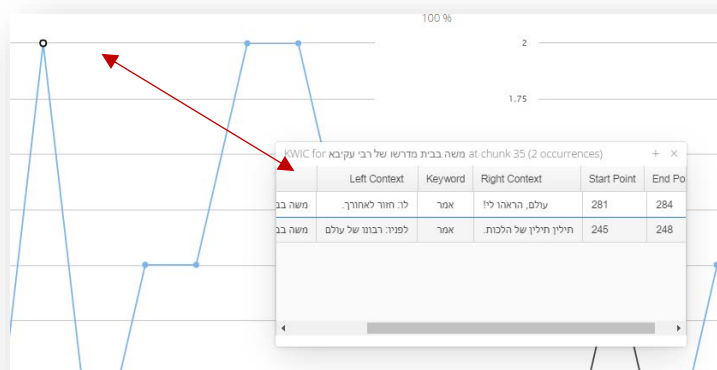
7.14. אפשר להוסיף לאותם גרפים מילים נוספות: נחזור למסך ה-Analyzer, ונסמן כעת את המילה 'משה'. לחיצה על כפתור גרף ההיקריות תצרף אותה לתמונה בצד המילה הקודמת.

7.15. יושם אל לב: סימון כמה מילים בו-זמנית, ולחיצה לאחר מכן על כפתור גרף ההיקריות, תצביע על נוכחות כל המילים הללו כקבוצה אחת בקו אחד על פני הגרף. על כן, כאשר ברצוננו להפריד בין מופעיהן של מילים שונות, עלינו להקפיד לבחור בכל פעם **מילה אחת בלבד** ולצרפה לגרף, באמצעות מעבר חוזר ונשנה בין מסך ה-Analyzer למסך ה-Visualizer. בשלבי מחקר ראשוניים, האפשרות לסימון בו-זמני עלולה להיראות מיותרת וחסרת תועלת, אולם עם התקדמות המחקר (וביחוד לאחר תיוג הטקסט או הקורפוס) נמצא שלעיתים יש ערך רב דווקא בבדיקת נוכחותם של כמה ממצאים יחד – אם כפנים שונות של אותה תופעה עצמה, ואם כדרך לבחון מחדש זיקה בין תופעות שהופרדו על ידינו בתהליך התיוג.

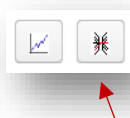
7.16. נוכל להוסיף מילים רבות אחרות באותה דרך – וגם להסיר אותן, כשברצוננו להתמקד בחלק מהממצאים בלבד. כדי להסיר מילים מהגרף יש ללחוץ על המילה המבוקשת בתחתית הגרף; המילה שעליה נלחץ תהפוך לאפורה, ותוסר מן הגרף עד שנחליט אחרת.



7.17. במידת הצורך, נוכל לעיין גם במסך זה בטבלת Kwic מקומית. לחיצה על כל אחת מהנקודות על פני הגרף תפתח טבלה קטנה על פני הגרף, ובה היקרויות המילה במקטע שסומן.

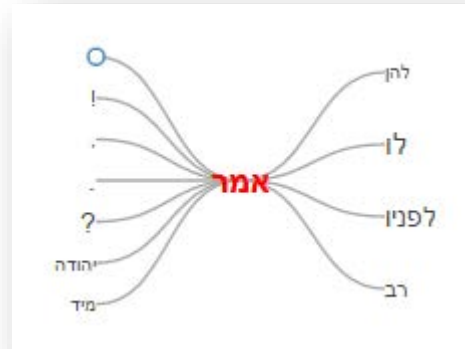


7.18. כזכור, CATMA מאפשרת גרף מסוג נוסף. אם נסמן מילה כלשהי ונלחץ על הכפתור השני שמתחת לצידה השמאלי של רשימת התוצאות, נקבל גרף 'עץ כפול' המציג את התנהגותה של מילה נתונה בכל הקשריה. המילה שביקשנו מופיעה כעת במרכז העץ; משני צידיה, המילים הסמוכות (סדר הקריאה – משמאל לימין, ולא כבשפה העברית); ככל שמילה גדולה יותר, כך היא נוטה יותר להופיע בסמיכות למילה הנבחרת.



.7.19

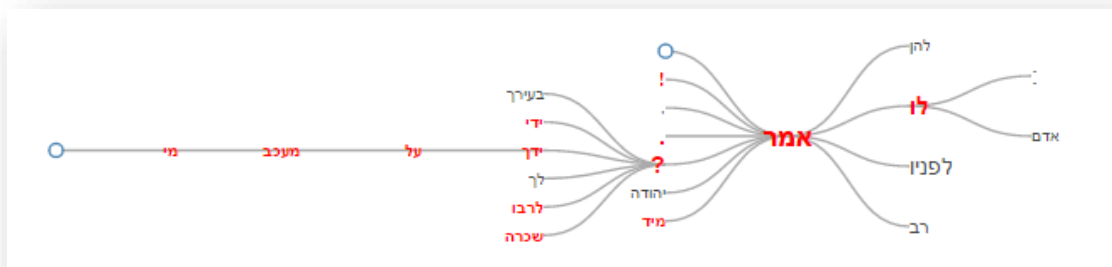
‘לפניו’ מופיעה יותר מן המילה ‘להן’, לאחר המילה ‘אמר’; ולפני מילה זו, סימן שאלה הוא האלמנט הנפוץ ביותר. תפקידו של גרף מסוג זה הוא לשקף בפנינו באופן ויזואלי את ‘התנהגותן’ ההקשרית של מילים מסוימות בטקסט, ואת זיקתן הישירה או העקיפה למילים שונות.



7.20. לחיצה על כל מילה בגרף העץ הכפול תאריך את ההקשרים להמשך המשפט, ותבליט את הצירופים האפשריים שלה בטקסט. כך למשל לחיצה על המילה 'מיד' בעץ דלעיל תבחר כי בהקשר זה, היא באה דווקא עם המילה 'לו' מצידה השני של המילה הנבחרת ('אמר'); ועוד היא תבחר מהן המילים הבאות לפניה (ובאמת, המשפט בטקסט המקורי הוא "...היה לך לעזרני. מיד אמר לו...").



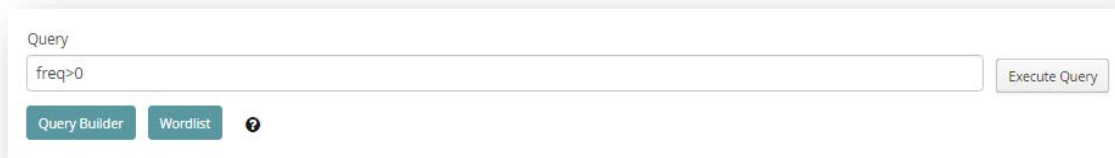
7.21. לחיצות נוספות על המילים השונות בגרף מסוג זה עשויות להניב לעתים עץ מסועף למדי, שיש בכוחו להבהיר בבת אחת מגוון צירופים אפשריים:



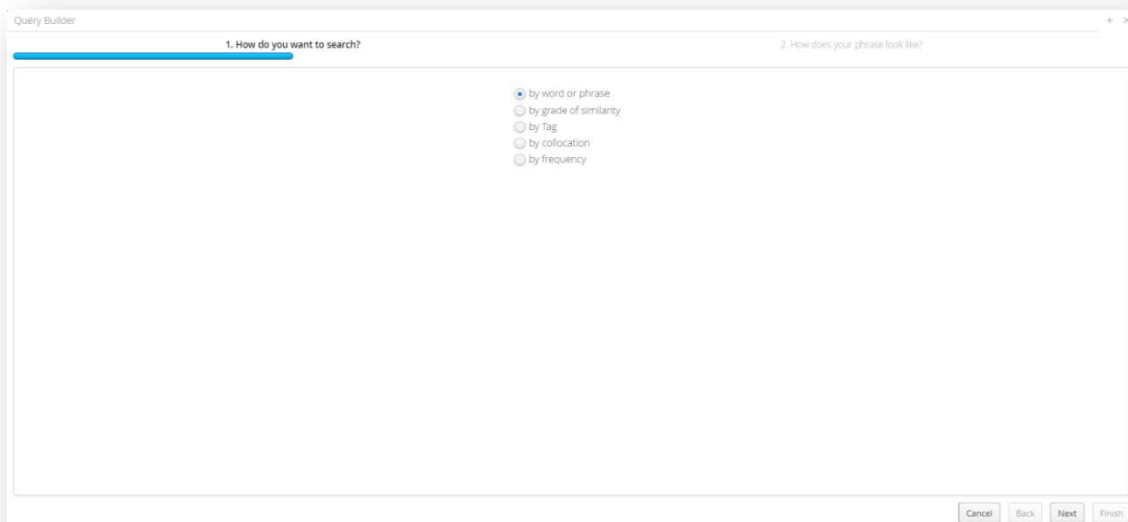
7.22. מטבע הדברים, ככל שהטקסט או הקורפוס עשיר ומורכב יותר, כך תועלתם של כלי הניתוח הללו עולה; אך גם השאילתא עצמה יכולה להיות עשירה ומורכבת יותר, אם מעט ואם הרבה, כפי שנראה מיד.

8. בונה השאילתות: היכרות ראשונית (שאילתות פשוטות)

8.1. נשוב כעת אל מסך ה-Analyzer על מנת לערוך היכרות ראשונית עם בונה השאילתות (Query Builder) של המערכת, הממוקם, כזכור, בצדו השמאלי-עליון של המסך. עד כה השתמשנו רק בשאילתא האוטומטית של רשימת המילים, אך CATMA מציעה בפנינו מגוון רחב ומשוכלל של שאילתות אחרות. בפרק הנוכחי נתמקד בשאילתות פשוטות, כלומר, כאלה הבנויות משלב חיפוש אחד בלבד.



8.2. השורה הריקה-למדי, שבה כתוב כעת רק $freq>0$ (מובנה של פקודה זו יובהר בהמשך), מאפשרת למשתמש המנוסה להקליד שאילתא בהתאם לכללי שפת החיפוש המיוחדת של התוכנה, ולתחביר המיוחד אותה; אנו נתעלם בינתיים מאפשרות זו (הערות אחדות לגביה תבואנה להלן), ונבחר במקומה ביצירת שאילתא מודרכת, הנשענת על 'שפה טבעית', שפת בני אדם. כך, לחיצה על כפתור Query Bulder תוביל אותנו אל המסך הבא:



8.3. בונה השאלות נועד להוליך אותנו צעד אחר צעד אל התוצאה המתבקשת, באמצעות שאלות המופנות אל המשתמש ומנחות את CATMA לפעול בהתאם לתשובותיו. לא נוכל להציגו במלואו, אך נתעכב על מרכיבי היסוד שלו ועל פונקציות שימושיות במיוחד.

8.4. בחלקו העליון של החלון נמצא תמיד קו כחול המסמן את השלב שבו אנו עומדים בתהליך יצירת השאלתא; כשהתהליך מושלם, הוא מגיע לקצה השני של החלון. מעליו מצוינת השאלה שהתוכנה מפנה אלינו בכל שלב. בחלקו התחתון של החלון נמצא תמיד את ארבע האפשרויות – Next, Back, Cancel ו-Finish (יושם אל לב: פקודות אלה נמצאות לעתים רחוק מהעין; משתמשים רבים שוכחים ללחוץ על Next או על Finish ואז אינם מבינים מדוע השאלתא אינה מתקדמת...).

8.5. השאלה הראשונה שנדרש לענות עליה היא 'כיצד – כלומר, על יסוד איזה פרמטר – ברצונך לחפש?' (How do you want to search?). בבואנו להשיב לה עומדות בפנינו חמש אפשרויות:

- חיפוש מונחה מילה או פסוקית (by word or phrase).
- חיפוש מונחה דומות (by grade of similarity).
- חיפוש מונחה תג (by tag).
- חיפוש מונחה צירוף (by collocation).
- חיפוש מונחה תדירות (by frequency).

8.6. הואיל ועדיין לא תייגנו דבר, נתעלם כעת מן האפשרות לחיפוש מונחה תג ונתמקד באפשרויות האחרות.

8.6.1. נבחר תחילה

בחיפוש מונחה

מילה או פסוקית,

ונלחץ על כפתור

Next. נקבל את

המסך הבא:

8.6.2. המערכת שואלת אותנו כיצד נראית הפסוקית שאנו מחפשים ('How does your phrase look like').

ומתמקדת תחילה במילה הראשונה. באפשרותנו להגדיר כיצד מילה זו מתחילה, מה היא מכילה, כיצד היא מסתיימת, ולחילופין – להגדיר מילה מדויקת. בהתחשב בדוגמה המובאת כאן, נבקש מהמערכת לאתר את כל המילים בקורפוס שמסתיימות בצירוף 'עלה'. לחיצה על כפתור Show in preview תראה את התוצאות 'עלה' ו'שעלה'.

8.6.3. אם ברצוננו למצוא צירופים המכילים יותר ממילה אחת, נלחץ על כפתור Add another word! (יושם

אל לב: במסכים קטנים כפתור זה נוטה 'להתחבא' באזור נסתר במורד החלון, ויש להגיע אליו באמצעות פס הגלילה המתאים מימין). כעת נוכל להגדיר את אחת המילים הבאות בצירוף: נוכל לקבוע שוב כיצד המילה הבאה מתחילה, מה היא מכילה, כיצד היא מסתיימת, ולחילופין, להגדירה בצורה מדויקת. ולא פחות חשוב: כעת נוכל גם להגדיר היכן נמצאת המילה השניה (The position of this word is...):
האם היא צמודה למילה הקודמת, או רחוקה ממנה (המערכת מאפשרת מרחק בשיעור של עד 10 מילים). במידת הצורך, נוכל כמובן להוסיף מילים אחרות לפסוקית באותה דרך.

8.6.4. בהמשך לדוגמא הנדונה כאן, נבקש כעת שהמילה השניה תסתיים בצירוף 'מרום', ונגדיר מרחק של שתי

מילים בין המילה המסתיימת ב'עלה' למילה המסתיימת ב'מרום'. לאחר לחיצה נוספת על Show in preview נקבל את התוצאה הבאה:

The screenshot shows the 'Query Builder' interface with two tabs: '1. How do you want to search?' and '2. How does your phrase look like?'. The first tab is active, showing search criteria for the first and next words. The second tab is also visible, showing the search results.

Search Criteria:

- The first word starts with: []
- The first word contains: []
- The first word ends with: [עלה]
- The first word is exactly: []
- The next word starts with: []
- The next word contains: []
- The next word ends with: [מרום]
- The next word is exactly: []
- The position of this word is: 2 words after the pre []

Buttons: Remove, add another word!

Results:

if your phrase contains more words you can: add another word!

Show in preview with a maximum total frequency of 50

Your search: wild="ile% %עלה" will match for example:

Phrase	Frequency
שעלה מרום למרום	2

8.6.5. יושם אל לב, כי באפשרותנו להשאיר חלק מההגדרות פתוחות לגמרי. כך למשל תיראה שאילתא שמבקשת למצוא את כל הפסוקיות בנות ארבע מילים, המסתיימות במילה 'אותיות' (לצורך העניין יכולנו גם להסתפק בתווים בודדים, ולא דווקא במילה שלמה נושאת משמעות): נשאיר את עמודת המילה הראשונה ריקה, נגדיר מרחק של ארבע מילים בין המילה הראשונה לשניה, ובעמודה השניה נבקש מילים המסתיימות ב-'אותיות' – ונקבל את הצירוף 'שיושב וקושר כתרים לאותיות' וכן את הצירוף 'שהיה קושר כתרים לאותיות' (כל אחד מהם מופיע אכן בסיפור אחר בקורפוס שלנו, ולהבדל ביניהם עשויה בהחלט להיות משמעות).

Query Builder

1. How do you want to search? 2. How does your phrase look like?

The first word starts with The next word starts with

The first word contains The next word contains

The first word ends with The next word ends with

The first word is exactly The next word is exactly

The position of this word is

If your phrase contains more words you can

with a maximum total frequency of

Your search will match for example:

Phrase	Frequency
שיושב וקושר כתרים לאותיות	1
שהיה קושר כתרים לאותיות	1

CATMA 8.6.6 מאפשרת לנו שאלות פתוחות אף יותר; למשל, נוכל לבקש מהתוכנה להציג בפנינו את כל הפסוקים המורכבות משלוש מילים, תהיינה אשר תהיינה. לשם כך נשאיר את שתי העמודות ריקות, ונגדיר רק את המרחק הרצוי:

Query Builder

1. How do you want to search?

The first word starts with	<input type="text"/>	The next word starts with	<input type="text"/>
The first word contains	<input type="text"/>	The next word contains	<input type="text"/>
The first word ends with	<input type="text"/>	The next word ends with	<input type="text"/>
The first word is exactly	<input type="text"/>	The next word is exactly	<input type="text"/>
		The position of this word is	3 words after the pre <input type="button" value="v"/>
		<input type="button" value="Remove"/>	

If your phrase contains more words you can

with a maximum total frequency of

Your search wild=" % % % "

Phrase	Frequency
▶ משה למרום מצאו	1
▶ כחו. כיון	1
▶ אמר לו	1
▶ ידי? אמר	1
▶ רבונו של עולם	5
▶ עלה במחשבה לפני	2
▶ אחרית חשש	1

יתרונה של שיטת חיפוש כזו טמון בכך שהיא מאפשרת חופש מוחלט בהגדרת הפסוקית המבוקשת – עד כדי ויתור על תכנים קונקרטיים ואפילו על גבולות תחביריים. יושם אל לב כי סימני פיסוק נחשבים ב-CATMA למילה, ולכן הצירוף 'ידי? אמר' שווה באורכו לצירוף 'משה למרום מצאו'. אם המשתמש מעוניין לוותר על סימני הפיסוק, עליו להתקין את הקובץ בהתאם לכך ולהזינו מחדש למערכת.

8.6.7. אין כמעט סוף לשילובים האפשריים. תהא אשר תהא בחירתנו, חשוב להקפיד ללחוץ על Show in

preview, משום שאם לא נעשה זאת המערכת לא תאפשר לנו ללחוץ על Finish כדי להשלים את בניית

השאלתא ולקבל את הממצאים בחלון התוצאות שבמסך ה-Analyzer.

8.7. בפרק הבא נלמד כיצד לבנות שאלתא מורכבת יותר, עשויה משלבים רבים. קוראים שאינם מעוניינים להעמיק

כעת את היכרותם עם בונה השאלות מוזמנים לדלג על שארית הפרק הנוכחי וכן על הפרק הבא כולו, על מנת

לעבור ישירות אל [חטיבת הפרקים המוקדשים לתיוג הפרשני](#).

8.8. נשוב אפוא אל הפרמטרים השונים העומדים ביסוד השאלות הפשוטות במערכת. כזכור, מלבד חיפוש מונחה

מילה או פסוקית (by word or phrase) עומדות בפנינו האפשרויות גם לחיפוש מונחה דומות (by grade

of similarity); חיפוש מונחה תג (by tag); חיפוש מונחה צירוף (by collocation); וחיפוש מונחה תדירות

(by frequency). נמשיך להתעלם לפי שעה מחיפוש מונחה תג, ובמקומו נסקור בקצרה את שלושת הסוגים

האחרים. כדי לתרגל כל אחד מהם, יש לחזור למסך ה-Analyzer וללחוץ שוב על כפתור ה-Query builder.

8.8.1. חיפוש מונחה דומות (by grade of similarity). נסמן אפשרות זו בחלון הראשון של בונה השאלות,

ונלחץ על Next. בחלון שנפתח לפנינו נוכל להזין מילה בתיבה הריקה שלמעלה על מנת לבקש מ-

CATMA למצוא לא רק אותה, אלא את כל המילים הדומות לה. שיעור הדומות נקבע באחוזים, ונשלט

על ידי המשתמש. כך למשל – אם להמשיך את דוגמת הקורפוס המודגם כאן – הצבת המילה 'אמר', קביעת

רמת הדומות על שמונים אחוז, ולחיצה על Show in preview תוליך לתוצאות 'אמר', 'אמרו', 'ואמר'.

8.8.1.1. לשיטת חיפוש זו חשיבות מיוחדת למחקר טקסטים בשפה העברית, משום שהיא מספקת פתרון

חלקי לבעיית השורש. אמנם, יש לתת את הדעת על קביעה מאוזנת של רמת הדומות בהתאם לצרכי

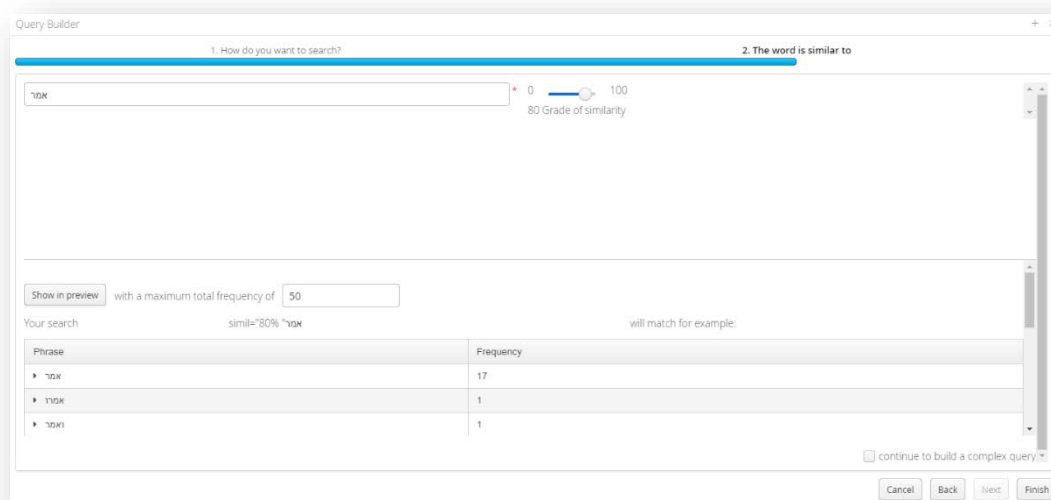
המחקר. האפשרות לראות תוצאות מוקדמות תוכל לסייע בכך; כך למשל שינוי רמת הדומות מ-80

אחוז ל-50 אחוז בדוגמא דלעיל תוסיף למילים הנזכרות גם את המילים 'כאשר' ו'למרום', שכמובן

אינן רלוונטיות לשורש המדובר.

8.8.2. תהא אשר תהא בחירתנו, בגמר התהליך נלחץ על Finish ונקבל את התוצאות הסופיות במסך ה-

Analyzer.



8.8.3. חיפוש מונחה צירוף (by collocation). בחירה באפשרות זו במסך הפתיחה של בונה השאלות, ולחיצה

על כפתור Next, תוליך אותנו לחלון חדש שבו נוכל לבקש מ- CATMA לחפש מילים בהתאם למידת הקרבה שלהן למילים אחרות בטקסט. בתיבה העליונה נקליד את המילה המבוקשת, בתיבה האמצעית את המילה שלידה המילה המבוקשת אמורה להיות (That appear near...), ובתיבה התחתונה את מספר המילים המירבי המפריד ביניהן (within a span of...). כך למשל חיפוש אחר כל הופעות המילה 'תורה' המתועדות במרחק של עד חמש מילים מהמילה 'אדם', יכלול את הופעת המילה 'תורה' במשפט 'יש לך אדם כזה ואתה נותן תורה על ידי?', אך ישמיט את הופעתה במשפט 'זו תורה וזו שכרה?'.

8.8.3.1. יושם אל לב: בניגוד לחיפוש מונחה מילה או פסוקית, שיטת חיפוש זו לא מניבה בתוצאותיה את

הצירופים עצמם, אלא משתמשת בהם כאמצעי לסינון הופעותיה של המילה המבוקשת; תוצאת השאלתא האמורה לא תהיה 'אדם כזה ואתה נותן תורה', אלא 'תורה' בלבד – אותה שמופיעה בסמיכות ל'אדם'. ואכן, אם נעביר לאחר מכן את התוצאה לטבלת ה- Kwic, נמצא כי מילת המפתח היא 'תורה'; שם נוכל לעיין במילים הסמוכות לה ככל שנבקש.

8.8.3.2. יושם אל לב עוד, כי האלמנט המילולי המבוקש, כמו גם האלמנט המשמש כאמצעי סינון, יכולים

להכיל יותר ממילה אחת: כך למשל אם נחפש את 'רבונו של עולם' הקרוב ל'הראהו לי' במרחק של עד 5 מילים, נקבל את הופעתה במשפט 'רבונו של עולם, הראהו לי', אך לא את הופעתה במשפט 'רבונו של עולם, זו תורה וזו שכרה'.

8.8.3.3. להשלמת בניית השאילתא נלחץ על Finish, ונקבל את התוצאות במסך ה-Analyzer.

8.8.4. חיפוש מונחה תדירות (by frequency). בחלון הפתיחה של בונה השאילתות נבחר אפשרות זו ונלחץ על

Next. בחלון הבא נוכל לבקש מ-CATMA לאתר עבורנו מילים שהמכנה המשותף היחיד להן הוא

תדירות הופעתן בטקסט או בקורפוס.

8.8.4.1. מידת התדירות מוגדרת בהתאם לשש האפשרויות הבאות:

(א) תדירות מדויקת (Exactly) – לחיפוש מילים המופיעות בדיוק X פעמים בטקסט או בקורפוס.

(ב) תדירות גדולה מ- (greaterThan) – לחיפוש מילים המופיעות אך ורק יותר מ-X פעמים

בטקסט או בקורפוס (לא כולל אלו המופיעות בדיוק X פעמים).

(ג) תדירות קטנה מ- (lessThan) – לחיפוש מילים המופיעות אך ורק פחות מ-X פעמים בטקסט

או בקורפוס (לא כולל אלו המופיעות בדיוק X פעמים).

(ד) תדירות גדולה מ- או שווה ל- (greaterOrEqualThan) – לחיפוש מילים המופיעות X

פעמים ומעלה בטקסט או בקורפוס.

(ה) תדירות קטנה מ- או שווה ל- (lessOrEqualThan) – לחיפוש מילים המופיעות X פעמים

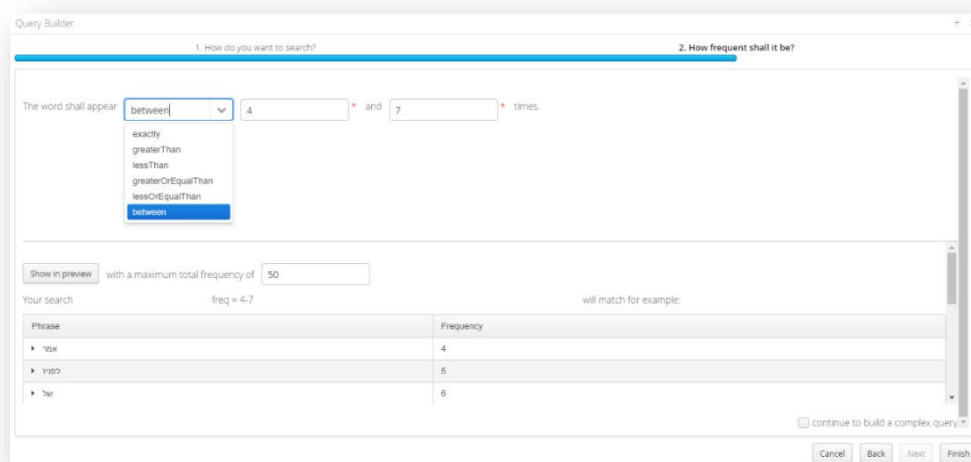
ומטה בטקסט או בקורפוס.

(ו) תדירות הנתונה בין X ל-Y (between) – לחיפוש מילים המופיעות X עד Y פעמים בטקסט

או בקורפוס (כולל אלו המופיעות בדיוק X או Y פעמים).

8.8.4.2. כך, בדוגמא הנדונה כאן, אם נבקש מהמערכת את כלל המילים המופיעות 4 עד 7 פעמים, נקבל

מילים כגון 'אמר' (4 פעמים), 'לפניו' (5 פעמים) ו'של' (6 פעמים).



8.8.4.3. לאחר הגדרת התדירות בהתאם לאחת מאפשרויות אלה נזכור ללחוץ על Show in preview,

ולאחר מכן על Finish. התוצאות הסופיות תופענה ברשימת הממצאים במסך ה-Analyzer.

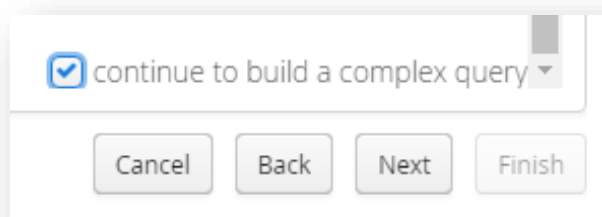
9. בונה השאילתות: שאילתות מורכבות

9.1. עד כה התמקדנו בשאילתות פשוטות בלבד – כלומר, כאלה העשויות מפעולת חיפוש אחת, תהא אשר תהא –

אך לא פעם נצפה מהמערכת לענות על שאילתות מורכבות יותר, המבוססות על שילוב פעולות חיפוש שונות

יחד. CATMA מאפשרת לנו להחליט בסוף כל שלב אם ברצוננו לסיים את החיפוש ולהביאו לידי גמר, או

לחילופין להמשיך לשלב נוסף.



9.2. לשם כך, נחזור שוב על אחת מפעולות

החיפוש שערכנו קודם, אך במקום

ללחוץ על כפתור Finish בגמר

התהליך, נסמן $\sqrt{\quad}$ בתיבה הקטנה שמעליו. כך נורה למערכת להמשיך לבנות שאילתא מורכבת יותר (continue (to build a complex query). הקו הכחול שבראש חלון בונה השאילתות יסוג כעת לאחור; האפשרות ללחוץ על Finish תיעלם אוטומטית באופן זמני, ובמקומה נוכל להתקדם לשלב הבא באמצעות לחיצה על Next.

9.3. כעת לפנינו חלון המעבר בין חלקי שאילתות מורכבות, ובו שאלה קבועה, המגדירה את היחסים בין השלב הקודם של השאילתא, לשלב הבא שלה: 'What do you want to do with the next query?.'

The screenshot shows a 'Query Builder' window with three tabs at the top. The third tab, '3. What do you want to do with the next query?', is active. Below the tabs, there are three radio button options: 'add more results' (selected), 'exclude hits from previous results', and 'refine previous results'.

9.4. כדי להגדיר את מטרת השלב הבא, נבחר אחת משלוש אפשרויות:

(א) **הוספת תוצאות לרשימה** (add more results). רשימת הממצאים הסופית תכלול את תוצאות השלב הראשון, וגם את תוצאות השלב השני.

(ב) **הוצא מרכיבים מן התוצאות הקודמות** (exclude hits from previous results). רשימת הממצאים הסופית תכלול את תוצאות השלב הראשון, למעט אלו מתוכן שנכללו גם בתוצאות השלב השני.

(ג) **זקק את התוצאות הקודמות** (refine previous results). רשימת הממצאים הסופית תכלול את תוצאות השלב הראשון, בתנאי שנכללו גם בתוצאות השלב השני.

9.5. בחירה בכל אחת מהן, ולחיצה על כפתור Next הנמצא בתחתית החלון, תוביל אותנו שוב לצעד הראשון בכל שאילתא ('How do you want to search?'), ולבחירה בין חיפוש מבוסס מילים, צירופים וכו'. הפעם, עלינו לזכור, איננו מתחילים לגמרי מחדש; כל ממצא יצטרף לממצאי השלב הקודם (או השלבים הקודמים) וישפיע על רשימת הממצאים הסופית. אולם האינטגרציה הסופית בין השלבים השונים לא תתגלה לעינינו בטרם השלמת השאילתא כולה. במילים אחרות, לחיצה על show in preview תיתן לנו בכל שלב תוצאות מוקדמות מקומיות, חלקיות, שאינן תלויות כלל בתוצאות השלבים האחרים, ואינן מושפעות מהן; רק לחיצה על Finish תספק את הסיכום הכולל של בקשתנו.

9.6. יושם אל לב, כי אין תלות הכרחית בין צורת החיפוש בשלבים השונים: אפשר לבנות שאילתא שתחילתה בחיפוש מבוסס מילים, המשכה בחיפוש מבוסס דומות, וסופה בחיפוש מבוסס תגים (שאליו עוד נגיע בהמשך). אין כמעט סוף לשילובים האפשריים, ואין כמעט סוף גם למספר השלבים. עם זאת, יש חשיבות רבה לסדר השלבים, המשפיע על קביעת היחסים ביניהם (כפי שעולה מתיאור אפשרויות ב' ו- ג' לעיל).

9.7. נחזור לשלוש האפשרויות בחלון המעבר בין השלבים. האפשרות הראשונה (add more results) היא הפשוטה ביותר: איננו נוגעים כאן בתוצאות שכבר הושגו בשלב הקודם של השאילתא, אלא רק מוסיפים עליהן תוצאות חדשות. אנו עשויים, דרך משל, לחפש בשלב ראשון את כל המילים הדומות ל'אמר', ובשלב השני את כל המילים המופיעות יותר משלוש פעמים, ומבקשים מהמערכת להציג בפנינו את תוצאות שני השלבים ברשימה כוללת אחת, מבלי שהאחת תגרע מרעותה או תשנה אותה. כל שעלינו לעשות לשם כך הוא לבחור continue to build a complex query בסוף השלב הראשון, לבחור add more results בחלון המעבר בין השלבים, וללחוץ על show in preview ועל Finish בסוף השלב השני.

9.8. שתי האפשרויות האחרות, לעומת זאת, משפיעות על התוצאות הקודמות בדרכים שתוארנה להלן, ולפיכך הן מעט מסובכות יותר.

9.9. ראשית, בבואנו לקבוע פרמטרים לסינון תוצאות או לעדכון הרשימה, עלינו לזכור שממצאי הפרמטרים החדשים עשויים שלא לחפוף לגמרי לאלה הישנים; עדכון התוצאות אינו מבוסס על עיון ברשימת התוצאות הקודמות ו'ניקוי' ידני של אלה מהן שאינן רצויות, אלא על חיפוש נוסף, שהצלבת ממצאיו עם אלו של השלב הקודם מעצבת את הרשימה מחדש.

9.10. כדי להבהיר זאת נעזר שוב בטקסטים שניתוחם מודגם כאן. אם נבנה שאילתא פשוטה שתכליתה לאתר את כל היקרויות הצירוף 'רבונו של עולם', נקבל חמשה ממצאים – כולם מאותו סיפור בקורפוס ('משה בבית מדרשו של רבי עקיבא'. בתמונה להלן אפשר לראות את רשימת הממצאים משמאל, ואת הקשריהם הטקסטואליים בטבלת ה- Kwic מימין (אגב, תמונה זו מבהירה היטב כיצד לחפש צירוף בקלות בלי עזרת בונה-השאילתות: די בציון הצירוף המבוקש ובהצבתו בין גרשיים בשורת הפקודה של ה- Query בחלקו השמאלי-עליון של המסך, ובלחיצה על Execute Query, על מנת לקבל את התוצאות המבוקשות).

Query

"רבות של עולם"

Query Builder

Wordlist

Execute Query

Documents and collections constraining this search

משה בבית מדרשו של רבי עקיבא

משה לא אמר שאלו לקבה

New Query

Result by phrase

Result by markup

Phrase	Frequency	Visible in Kwic
רבות של עולם	5	<input checked="" type="checkbox"/>
משה בבית מדרשו של רבי עקיבא	5	

Document/Collection	Left Context	Keyword	Right Context	Start Pos	End Point
משה בבית מדרשו של רבי עקיבא	... וזו מעבד על ידך	רבות של עולם	למנוחות אמר לפניו	100	113
משה בבית מדרשו של רבי עקיבא	... הראונו לל אמר	רבות של עולם	הקנות אמר לפניו	256	269
משה בבית מדרשו של רבי עקיבא	... יש לך אדם מה	רבות של עולם	הוא אמר לפניו	495	508
משה בבית מדרשו של רבי עקיבא	... הראינו חורונו הראוי	רבות של עולם	לפני אמר לפניו	592	605
משה בבית מדרשו של רבי עקיבא	... זו חורונו שברת	רבות של עולם	במקומו אמר לפניו	705	718

9.11. אך מה אם ברצוננו לבנות רשימת תוצאות הכוללת אך ורק את אותם מופעים של 'רבותנו של עולם' שאינם סמוכים למילה 'אדם'? לשם כך נחזור על השלב הראשון של השאילתא בכונה-השאילתות (כלומר, נחפש by word or phrase את כל הרצפים המתחילים ב'רבותנו' ומסתיימים ב'עולם', במרחק שתי מילים), ובמקום לסיימה נורה למערכת להמשיך לבנות שאילתא מורכבת.

9.12. בחלון המעבר נבקש להסיר תוצאות מן השלב הקודם – exclude hits from previous results – ומיד נתבקש להגדיר את מידת החפיפה בין תוצאות שני השלבים: האם צירוף שניהם מותנה:

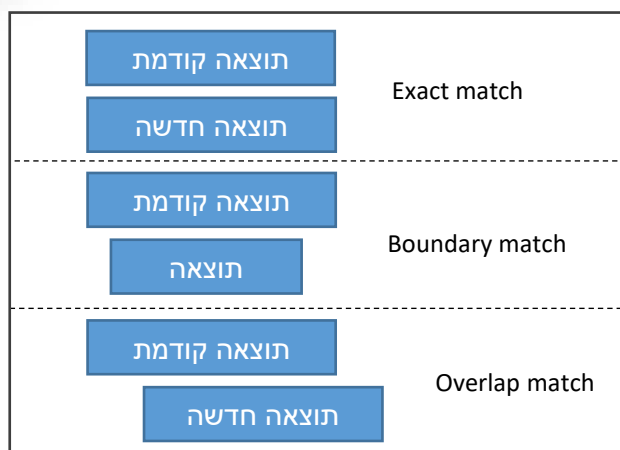
☐ add more results
☒ exclude hits from previous results
☐ refine previous results

Please choose what you consider a match:

exact match
 exact match
 boundary match
 overlap match

- (א) בהתאמה מדויקת של גבולותיהם (exact match).
 (ב) בגבולות משותפים (boundary match), במקרה שתוצאות השלב השני קצרות מתוצאות השלב הראשון.
 (ג) בחפיפה חלקית בלבד (overlap match), כלומר, כזו שאינה נכללת בהכרח בגבולות משותפים.

9.13. לעת עתה נבחר בחפיפה חלקית בלבד (overlap match), ונלחץ על כפתור Next.



9.14. בשלב הבא נתבקש להגדיר את הפרמטר הרלוונטי לשלב החיפוש השני, ובהתאם לרצוננו נבחר הפעם בחיפוש

by collocation. נלחץ שוב על כפתור Next.

9.15. בחלון הבא נבקש מהמערכת לאתר את כל מופעי המילה 'עולם' הסמוכים למילה 'אדם' במרחק של עד 5 מילים.

CATMA תמצא מופע אחד כזה:

עולם *

that appear near *

אדם

within a span of *

5

Note: the preview frequency limit will be applied to each sub result because of this no preview results doesn't necessarily mean that there will be no query results at all!

Show in preview with a maximum total frequency of 50

Your search "עולם" & "אדם" 5

Phrase	Frequency
עולם	1

9.16. לאחר לחיצה על כפתור Finish נקבל ארבעה ממצאים במסך ה-Analyzer; מתוך חמשת מופעי 'רבונו של

עולם' הוסר זה הסמוך למילה 'אדם':

Query: (wild="5" "אדם" & "עולם") overlap

Execute Query

Documents and collections constraining this search

- משה בבית מדרשו של רבי עקיבא
- משה לא אמר שלום לקב"ה

Result by phrase

Phrase	Frequency	Visible in Kwic
רבונו של עולם	4	<input checked="" type="checkbox"/>

Result by markup

Document/Collection	Left Context	Keyword	Right Context	Start Pos	End Point
משה בבית מדרשו של רבי עקיבא	מי מעב על דרך	רבונו של עולם	לאחרות אמר למיני	100	113
משה בבית מדרשו של רבי עקיבא	הראובן ליל אמו	רבונו של עולם	הלכות אמר למיני	286	289
משה בבית מדרשו של רבי עקיבא	הראובני חורח, הראובני	רבונו של עולם	למיני אמר למיני	592	605
משה בבית מדרשו של רבי עקיבא	זו חזרה וזו שכרה	רבונו של עולם	במקורן אמר למיני	705	718

9.17. יושם אל לב: בחרנו בחפיפה חלקית בין התוצאות (overlap), משום ששני השלבים חולקים יחד אך ורק את

המילה 'עולם'. אילו בחרנו אחרת, כל חמש התוצאות היו נשארות ברשימת הממצאים והשלב השני לא היה

משפיע עליהן כלל.

9.18. אך מה אם נבקש את ההיפך – כלומר, נבקש

להפיק רשימה הכוללת אך ורק את אותם מופעים

של 'רבנו של עולם' הסמוכים למילה 'אדם'?

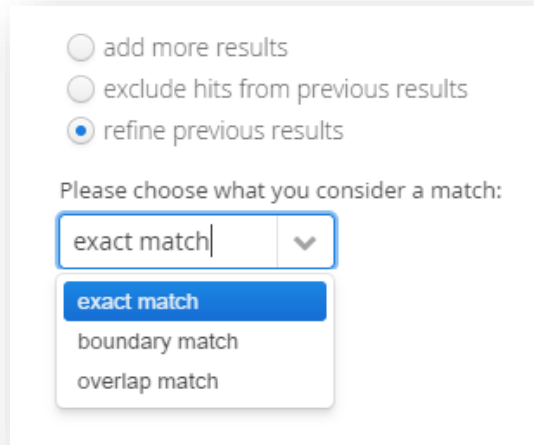
לשם כך נחזור תחילה שוב על השלב הראשון.

לאחר מכן נבקש מהמערכת להמשיך לבנות

שאלתא מורכבת, ובחלון המעבר בין שלבי

השאלתא נבחר באפשרות refine previous

results ושוב בהתאמה בדרגת overlap.

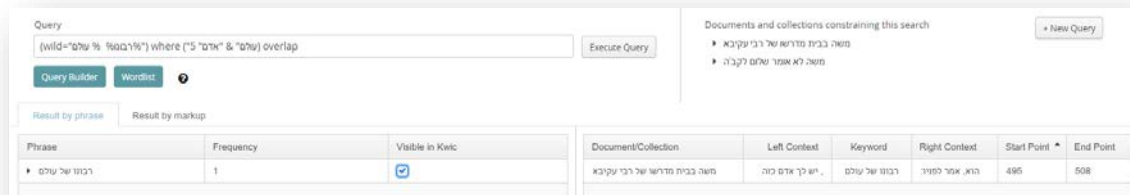


9.19. בחלון הבא נבחר שוב חיפוש by collocation, ונחזור על אותן הגדרות שתיארנו לעיל (כלומר, נורה למחשב:

מצא את כל מופעי המילה 'עולם' הסמוכים למילה 'אדם' במרחק של עד חמש מילים). לאחר לחיצה על Finish

נקבל ברשימה שבמסך ה-Analyzer ממצא אחד בלבד. המחשב סינן את ארבעת הממצאים שלא התאימו

להגדרות בשלב השני:



Phrase	Frequency	Visible in Kwic
רבנו של עולם	1	<input checked="" type="checkbox"/>

Document/Collection	Left Context	Keyword	Right Context	Start Point *	End Point
משה בבית מדרשו של רבי עקיבא	יש לך אדם כזה.	רבנו של עולם	הוא אמר לפניו	495	508

9.20. בדרך דומה אפשר לנסח שאלות מורכבות מגוונות למדי, ויהיו אשר יהיו צירופיה ותנאיה. הקפדה על סדר

פעולות נכון, ועל הגדרה מספקת של היחסים בין שלבי החיפוש, חיונית להפקת תוצאות רצויות.

9.21. אמנם בנקודה מסוימת נגלה כי בונה השאלות, נוח ככל שיהא, אינו עונה על כל צרכינו. בשלבים מתקדמים של

העבודה, נמצא שברצוננו לשנות מעט את אחד ממרכיבי השאלתא, או פשוט לנסח אותה בעצמנו מהחל ועד

כלה – אם משום שיצירתה באופן מודרך נראית לנו מסורבלת ואיטית, ואם משום שהשאלתא המסוימת

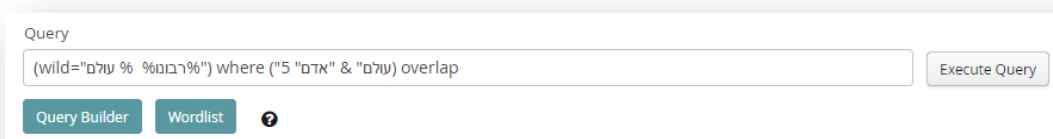
שמעסיקה אותנו חורגת מגבולות האפשרויות המוצעות בבונה השאלות. לשם כך מציעה המערכת את תיבת

ניסוח השאלות בחלקו השמאלי-עליון של מסך ה-Analyzer.

9.21.1. שפת השאילתות אינה אינטואיטיבית לגמרי, אך לפורמולציה שלה יש הגיון מעין-מתמטי: פעולות קבוצות מיוצגות במילים קבוצות כגון where, בסימני חשבון מוסכמים כגון -, <, ועוד; שלבי-ביניים נתונים בתוך סוגריים, הקובעים הן את ההירארכיה הפנימית שלהם בכלל מרכיבי השאילתא, והן את סדר הפעולות הרצוי (בדיוק כמו במשפטים מתמטיים, המערכת ממלאת תחילה את ההוראות שבסוגריים, ואז משלבת את התוצאה בהוראות המופיעות מחוץ לסוגריים).

9.21.2. CATMA לא מסתירה את שפתה מאחורי הקלעים של בונה השאילתות הידידותי-למשתמש, אלא מציגה אותה עם השלמת התהליך בתיבת הניסוח, מעל התוצאות. לפיכך, הדרך הפשוטה והיעילה ביותר ללימוד שפת השאילתות היא לעקוב בכל פעם אחר הכיתוב המופיע בתיבת הניסוח, לאחר השלמת שאילתא בבונה השאילתות. עד מהרה ילמד החוקר את מרכיבי השפה, ויגלה כיצד הוא יכול 'לשחק' בהם מעט או הרבה על מנת לרתום אותם לצרכיו. די בשינוי קל של שורת הפקודה ובלחיצה על כפתור Execute Query הנמצא מימינה על מנת לערוך ניסויים כאלה. יתר על כן, תשומת לב ללשון הפקודה בתיבה זו עשויה לסייע לחוקר להבין טוב יותר את השאילתות שבנה, ולשפרן.

9.21.3. כך למשל מבט נוסף על שורת הפקודות בתמונת המסך של השאילתא האחרונה לעיל מלמד כי המערכת בנתה שאילתא מורכבת משני שלבים – באחד מהם זיקה כלשהי בין המילים 'רבונו' ו'עולם', ובשני זיקה אחרת בין המילים 'עולם' ו'אדם' (נניח לעת עתה לפרטיה המדויקים של כל זיקה) – כאשר השלב השני הוגדר כתנאי מחייב לברירת ממצאי השלב הראשון (כפי שעולה מהמילה where הנתונה ביניהם), וככל שהללו חופפים (כפי שעולה מהמילה overlap המופיעה לאחר השלב השני).



9.21.4. קוראים המעוניינים להעמיק את ידיעותיהם בשפת השאילתות של CATMA מוזמנים לעיין בנספח המצורף למדריך זה בסופו. ואולם, הואיל והדרך הטובה ביותר ללימודה הוא בהתנסות מדורגת, והואיל ועיקר תרומתה של השאילתא המורכבת עתידה לבוא לידי ביטוי בניתוח טקסטים מתוייגים, נעבור כעת לתיאור גולת הכותרת של המערכת – כלי התיוג. לכך יוקדשו הפרקים הבאים.

10.1. הפרקים הקודמים במדריך הוקדשו לצדדים שונים של ניתוח טקסטואלי אוטומטי. מושא הבדיקה היה הטקסט או הקורפוס עצמו – שכבת ה- data 'אובייקטיבית'; פעולת הבדיקה מומשה כמעט במלואה על ידי המחשב, שעשה ממילא את מה שהוא יודע לעשות הכי טוב, או, על כל פנים, את מה שהוא יודע לעשות טוב יותר מאיתנו: לספור. תפקיד החוקר, בשלבים אלו, לא היה אלא להנחות את המחשב אל הספירה הנכונה, המועילה לו ביותר.

10.2. בהתחשב בקורפוס הזעיר המודגם כאן, ספק אם ספירה זו הועילה במשהו; יש לקוות אמנם שהיה בה כדי לגרות את הדמיון באשר לתוצאותיהן של ספירות מורכבות של טקסטים בהיקף גדול יותר. הפרקים הבאים, לעומת זאת, מציגים את הפנים האחרות של מערכת CATMA – כלומר, את תרומתה להעמדת פרשנות אנושית סובייקטיבית, ובהמשך גם לאופרציונליזציה חישובית שלה.

10.3. תחילה עלינו להבהיר את מושג המפתח של הפרק הנוכחי – תיוג (tagging), או ביאור, או אנוטציה (annotation) – ולעמוד על החידוש הטמון בו, משעה שמעניקים לו משמעות רחבה. למושג זה תפקיד מרכזי במדעי הרוח הדיגיטליים, אך על פי רוב הוא נתפס כאמצעי פונקציונלי שתכליתו לתאר את הטקסט הנחקר: להבהיר היכן הכותרת, היכן גוף הטקסט, היכן מופיעים ציטוטים, ועוד. כך נוצרת שכבת מטא-דאטה מינימליסטית המשקפת למחשב עצמו כיצד אנו קוראים את הטקסט. מדובר – אם לפשט את הטיעון – במידע 'אובייקטיבי', או לפחות מוסכם, על אודות כמה מתכונות רצף המילים והתווים שהוזן למחשב. משמעות דומה של המושג מוכרת ודאי למשתמשי הרשתות החברתיות, כאמצעי לציון זהות הדמויות בתמונה נתונה; גם בהקשר זה מדובר בשכבת מטא-דאטה מינימליסטית, המוסיפה רובד של מידע 'אובייקטיבי', או לפחות מוסכם, לאובייקט גרפי – ולא רק כדי לשקף למחשב כיצד 'לפענח' את התמונה, אלא גם כדי לייצר שפה משותפת בין בני אדם, משתמשי-המחשב, בכל הנוגע לפרטי אותה תמונה.

10.4. אולם CATMA מעודדת אותנו לתפיסה אחרת של המושג. אם 'תיוג', כפשוטו, הוא 'הוספת תג' על גבי אלמנט קיים, מדוע להסתפק בתג אובייקטיבי ומוסכם? מדוע להסתפק בהוספת שכבת מידע צרה, כאשר אפשר להעמיס

על הטקסט את כל מה שראוי להיקרא 'פרשנות' או 'קריאה' – בהקשריהם הרחבים, ואפילו הסובייקטיביים ביותר האפשריים? האם אין פרשנותנו ראויה להיחשב ל'מטא-דאטה' מסדר חדש?

10.5. אם לנסח את הדברים באופן אחר, הרי כשם שהמחשב איננו יודע שרצף תווים מסוים הוא 'כותרת', כך גם אין הוא יודע שרצף תווים אחר הוא 'מטאפורה'; כך גם אין הוא יודע להבחין בין שימוש במילים-כפשוטן, לבין שימוש אירוני; כך גם אין הוא יודע לזהות טקסט כחוקתי, פילוסופי או נראטיבי, כריאליסטי או פנטסטי, וכל כיוצא בזה. אבל אנחנו יודעים; כך לפחות אנו מדמים לחשוב. התיוג הפרשני, כפי שיוצג מיד, הוא דרך לתרגם את הידיעות הללו לסימנים על הדף, או על מסך המחשב, באופן המעניק לנו תפיסה חדשה של הטקסט העומד לניתוח.

10.6. מבחינה מתודולוגית, CATMA מציעה לנו דרך יעילה מאוד להותיר בטקסט, או לפחות על גביו, חותם של עמלנו הפרשני, ובכך הופכת אותו מ'תורה שבעל פה' ל'תורה שבכתב': בהיותו מתועד, אנו יכולים לשקף אותו לעצמנו ולקהלנו; אנו יכולים לפתח אותו באופן שיטתי, לעקוב אחריו, לבחון אותו במבט לאחור, לתקן אותו במידת הצורך; אנו יכולים להותיר אותו ב'צריך עיון', להשאירו כעמדה מסופקת, כהשערה או כהיסוס; אנו יכולים להשוות אותו לפרשנותם של אחרים; וחשוב מכל – אנו יכולים לאמוד את יחסי הגומלין שלו עם מושאו, עם הטקסט המתפרש, ולדון בהם. בה בעת עלינו לזכור כי בניגוד לסכימות תיוג נפוצות, תיוג פרשני חופשי מן הסוג המוצע כאן אינו חותר לסטנדרטיזציה של שכבת ה-metadata, אלא דווקא ליצירתיות ולגמישות מלאה.

10.7. תיוג פרשני, אם כן, הוא דרך להפוך את האינטואיציות של החוקר לגוף-ידע מתועד, מדיד, זמין לניתוח נוסף ולעיבוד חוזר – וממילא, לאופרציונליזציה חישובית של שאלת מחקר מסורתית; הוא דרך לצלוח את הפער הסמנטי – ולא רק את הפער הסמנטי היסודי שבין תווים לטקסט, אלא גם את זה שבין טקסט ומשמעות, ואפילו במובנה הרחב ביותר. בנקודה זו כבר ברור למדי כי איננו מדברים עוד במחקר 'אוטומטי', ואפילו לא במחקר שהמחשב הוא המחולל המרכזי שלו. תפקידה של הסביבה הדיגיטלית המתוארת כאן, אפוא, הוא לספק תמיכה רבת עוצמה לפרוצדורות מחקריות כמעט-רגילות, מוכרות היטב: אין זו קריאה רחוקה העלולה לאבד קשר עם הטקסט עצמו, אלא להיפך; זו היא קריאה קרובה המחויבת למגע רציף עם מושא עיונה.

10.8. ארבעת הפרקים הבאים יוקדשו להצגת התיוג הפרשני משלב הכנת מערכת התגים ומסך התיוג, דרך התנסות בתיוג ידני ואף בתיוג אוטומטי-למחצה. לאחר מכן נשוב אל מסך ה-Analyzer – אך הפעם, מלבד הטקסט הגולמי, תהא ברשותנו שכבה פרשנית עשירה, שניתוחה החישובי עשוי להתוות כיווני מחקר חדשים.

11.1. מערכות התגים ב-CATMA מאפשרות גמישות מירבית בקביעת יעדי המחקר. המשתמש רשאי להגדירן כמעט בכל דרך העולה על דעתו; הוא יכול לייחס להן משמעות קונקרטית צרה, או להעמיס עליהן מבנים של משמעות עודפת; הוא יכול לפצל אותן ל'ספריות' תגים שונות, או להעמידן במבנה הירארכי מרובד; הוא יכול לעבור ביניהן בנקל, ואף לשנות אותן תוך כדי עבודה.

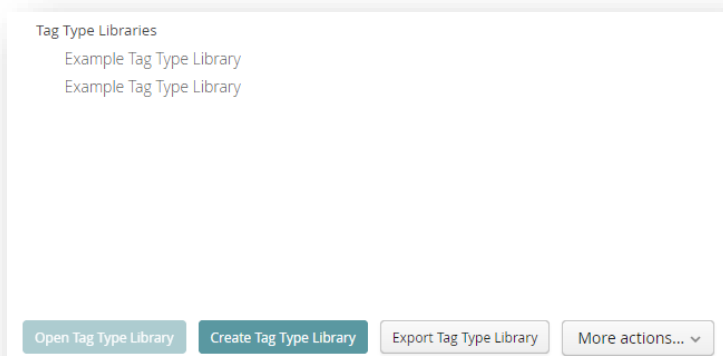
11.2. עקרון יסוד במנגנון התיוג של CATMA טמון בהפרדה המלאה בין (א) הטקסט, כלומר, שכבת ה-data שלה מוקדש מחקרנו, (ב) מערכת התגים, כלומר, סכימה מופשטת של קטגוריות פרשניות שלאורן נקרא את הטקסט,



11.3. נבהיר זאת בלשון אחרת: גם אם נוצרה עבור טקסט מסוים אחד, מערכת התגים מתקיימת (כמעט) לעולם בנפרד; גיבוש המערכת, כשלעצמו, אינו חופף ליישומה על טקסט נתון, גם אם הוא מושפע ממנו. לפיכך, כאשר אנו מגבשים מערכת תגים אנו מכינים כלי עבודה שיוכל לשמש בניתוח טקסטים רבים, או לא לשמש כלל ועיקר. אשר על כן, כאשר אנו משתמשים בה לתיוג טקסט מסוים, איננו משפיעים בהכרח לא עליה ולא על הטקסט, אלא מייצרים שכבת מטא-דאטה נוספת, שכבת-ביניים כביכול, שבה נשמרים תיוגינו השונים, או אזור נפרד שיכונה בעברית, כאמור, 'אסופת התיוג' (markup-collection).

11.4. הודות למבנה המודולארי של המערכת נוכל לפתוח שוב בהזדמנות אחרת את אותו טקסט מחדש, ולהעמידו להתבוננות נוספת מזווית מבט אחרת לגמרי; הודות למבנה המודולארי נוכל לאמץ את מערכת התגים בשעה שנטפל בקורפוס שונה בתכלית. וכשם שבאפשרותנו ליישם את אותה מערכת תגים על טקסטים רבים, כך באפשרותנו לעצב מערכות תגים מרובות עבור הטקסט האחד, ואפילו אסופות תיוג מרובות; הכל בהתאם לצרכי המחקר. ההפרדה המוחלטת בין השכבות השונות מבלבלת לעתים את המשתמש המתחיל, המתקשה להבין מדוע אינו יכול 'פשוט לתייג וזהו'; אך עד מהרה עתידים להתגלות היתרונות המעשיים והקונספטואליים שלה.

11.5. השלב הראשון בהכנת מערכת תגים נמצא בתחתית מסך ה- Repository Manager, בצידו השמאלי. שם,



מתחת לטבלת הקורפוסים,

נמצא את המקום שבו

נשמרות ומנהלות ספריות

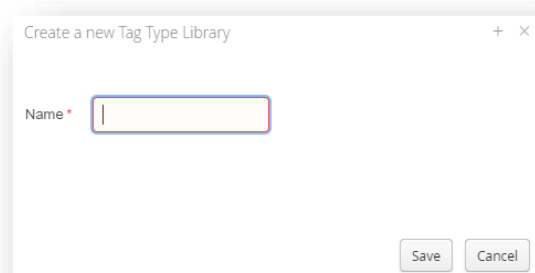
התגים (Tag Type)

(Library). כדי ליצור

ספריית תגים חדשה נלחץ

על כפתור Create Tag

Type Library.



11.6. כעת נפתח חלון חדש שבו נתבקש לתת שם

לספרייה החדשה. נבחר בשם מתאים – למשל,

שם ההולם את ייעודם הכולל של התגים

המתוכננים, אך לא דווקא בזיקה לטקסט או

לקורפוס ספציפי – ונלחץ על save.

11.7. יושם אל לב: 'ספריית' תגים אינה תג בפני עצמו, ואף אין היא חופפת למערכת התגים. כשמה כן היא: ספרייה,

כלומר, מעין ארון שבתוכו מונחים ספרים (או אובייקטים אחרים). הארכיטקטורה המבנית של CATMA קובעת

כי –

11.7.1. כל ספריית תגים (tag library) מכילה מערכות תגים (tag-sets);

11.7.2. כל מערכת תגים עשויה להכיל תגים (tags או tag types), שעשויים להכיל תגי-משנה (sub-tags);

11.7.3. כל תג או תג-משנה עשוי להתמייך בהתאם לתכונות ספציפיות (properties).

11.8. אף על פי שלכל אחת מן הרמות יש כותרת עצמאית, אין לבלבל את שתי הרמות הראשונות עם שלוש הרמות

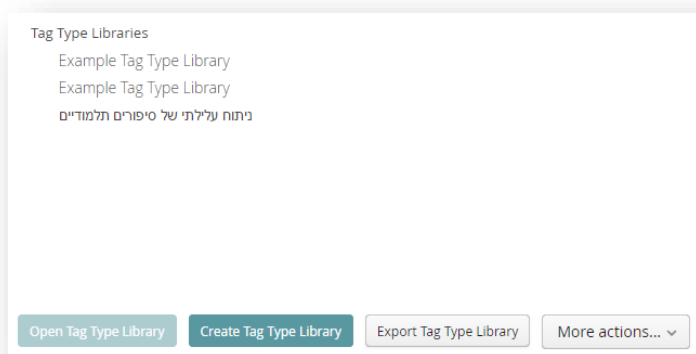
האחרונות במבנה הירארכי זה: ספריות תגים ומערכות תגים – אינן 'תגים' בפני עצמן, כי אם מיכלים לתגים; אי

אפשר לסמן אותן על גבי הטקסט, אלא רק את מה שבתוכן, כלומר: תגים, תגי-משנה ותכונות. נלמד להלן כיצד

ליצור כל אחת מן הרמות.

11.9. בהתחשב בדוגמה הנדונה כאן, נקרא לספריית התגים 'ניתוח עלילתי של סיפורים תלמודיים' ונלחץ על save.

הספרייה החדשה נוספה לספריות האחרות, ומוצגת כעת במסך ה-Repository manager. ואולם, עדיין איננו יכולים לתייג; עלינו לפתוח את הארון (כלומר, את הספרייה), לקבוע בתוכו מגירות (אלו הן מערכות התגים, כפי שנראה מיד), ורק אז להכניס לתוכן את כלי העבודה הדרושים לנו, ולארגן אותם באופן מועיל.



11.10. נסמן את הספרייה

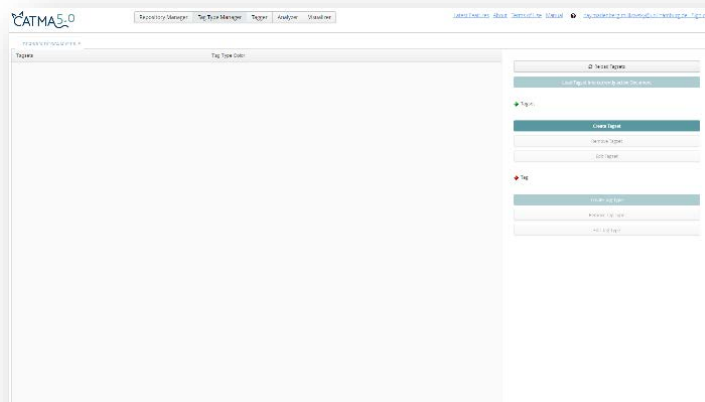
החדשה ונלחץ על כפתור

Open Tag Type

Library.

11.11. בחירה זו מובילה

אותנו לראשונה למסך ה-Tag Type Manager. בשלב זה רובו לבן, אפור וריק, משום שספריית התגים שלנו ריקה.



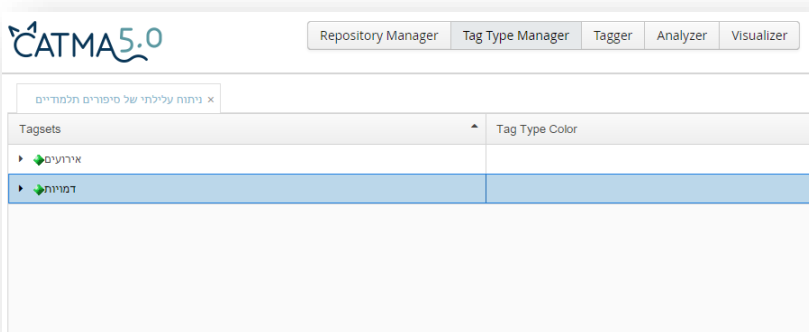
11.12. כעת נוכל להגדיר מערכות תגים – אחת או יותר. יוזכר שוב: מערכות התגים, כשלעצמן, אינן כלי התיוג,

אלא מעין 'ספרייה פנימית', ארגז או מגירה ייעודית, שבתוכה נרכז כלי עבודה בעלי מכנה משותף כלשהו; הן עדיין לא יכולות לשמש לאפיון התופעות שאנו מבקשים לתייג, אלא רק לקביעת המכנה המשותף של תופעות אלו. לפיכך אין טעם להגדיר אותן באופן צר מדי (שגיאה בהבנת מקומן בהירארכיה של התיוג ב-CATMA עלולה לשבש מאוד את המשך העבודה, והיא מן הטעויות הנפוצות של המשתמש המתחיל).

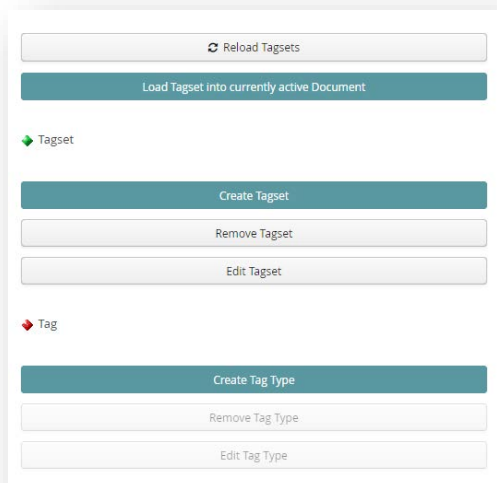


11.13. נגדיר מערכת תגים באמצעות לחיצה על כפתור Create Tagset המופיע בצידו הימני של המסך. כעת יפתח חלון חדש, ובו נתבקש לתת שם למערכת התגים. לאחר בחירת השם נלחץ על כפתור save. המערכת החדשה תופיע בטבלת ה- Tagsets בצידו השמאלי של המסך.

11.14. כדי לפשט את הדוגמא, נתרכז בשני היבטים בלבד של מחקר העלילה: אירועים ודמויות. לפיכך ניצור מערכת תגים אחת בשם 'אירועים', ומערכת תגים נוספת בשם 'דמויות'. יושם אל לב: מערכות התגים לעולם



מקבילות זו לזו בהירארכיה של המערכת, ואינן כפופות אחת לשנייה או נובעות אחת מהשנייה.



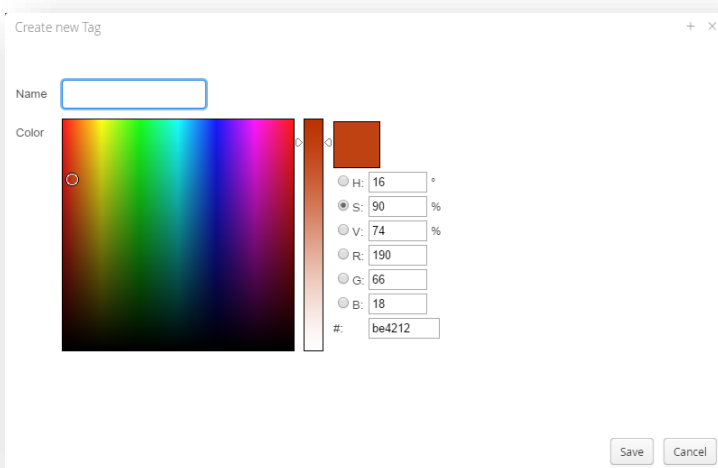
11.15. כעת נוכל סופסוף ליצור את כלי העבודה ששימשו אותנו בהמשך – את התגים. לשם כך נסמן את מערכת התגים שבתוכה נבקש למקם את התגים המיועדים, ונלחץ על כפתור Create Tag Type הנמצא בחלקו הימני-תחתון של המסך.

11.16. בחלון החדש שנפתח נתבקש לתת שם לתג, ולבחור את הצבע שבו יסומן על גבי הטקסט. בכל פעם שחלון כזה נפתח המערכת מציעה באופן אוטומטי צבע אחר, אך המשתמש רשאי לקבוע את

הצבע כרצונו באמצעות הזזת העיגול הקטן הנמצא על מפת הצבעים הריבועית, וכן באמצעות הזזת החיצים המשולשים הנמצאים לאורך העמודה האנכית הקובעת את מידת בהירותו של הצבע (בחירת הצבע וקביעת שמו

של התג יכולה להשתנות גם בהמשך, במקרה הצורך). הצבע הנבחר מופיע בריבוע הקטן שבחלקו הימני-עליון של החלון.

11.17. כאשר עוסקים בתופעות שיש ביניהן מדרג, עשויה להימצא תועלת בקביעת גוונים שונים של צבע אחד – גוון שונה לכל תג; כאשר עוסקים, לעומת זאת, בתופעות שיש ביניהן חץ של ממש, עשויה להימצא תועלת



בקביעת צבעים שונים לגמרי. קשה אולי להבין את תועלתן המחקרית של בחירות אלה כעת, אך היא עתידה להתגלות בהמשך, כאשר נבקש לקבל התרשמות ראשונית מטקסט שכבר תויג: אם הצבעים השונים נקבעו בהתאם למבנה הגיוני כלשהו,

יקל עלינו לתפוס מגמות מסוימות בטקסט כבר בפעולת התבוננות ראשונית בצבעים המסומנים על גביו.

11.18. בהתחשב בדוגמא המוצעת כאן, נעצב תחילה תגים תחת מערכת התגים 'דמויות'. ראשית נבקש להבחין בין 'מספרים' ל'דמויות פעילות בסיפור', ונייצר תג שונה לכל אחד מהם.

11.19. לאחר מכן נבקש להבחין בין סוגים שונים של 'דמויות פעילות בסיפור'; לשם כך נסמן את התג הרלוונטי, ונלחץ שוב על Create Tag Type (כפתור זה משמש הן ליצירת תגים, הן ליצירת תגי-משנה). נוסיף כעת בזה אחר זה את תגי-המשנה הבאים: 'דמות מקראית', 'דמות מעולמם של חז"ל', 'דמות אנונימית', 'דמות אלוהית'.

11.20. כעת נטפל במערכת התגים 'אירועים': נסמן אותה – על מנת שהתגים החדשים יוכפפו לה – ונלחץ על Create Tag Type. נוסיף את התגים 'אירועים מקדמי עלילה' ו'אירועים בעלי משמעות מיוחדת'. כדי לא להיכנס לפרטים ספציפיים יתר על המידה, נאמר רק שהאבחנה המוצעת כאן (אמנם, באופן פשטני למדי), נשענת על אבחנה מקובלת בתיאוריה של הנראטיב בין סוגים שונים של אירוע (event): כל שינוי מצב בעולם המיוצג בסיפור, ולו גם הקטן והמינורי ביותר, יכול להיחשב ל'אירוע'; אלו הם, בלשוננו כאן, 'אירועים מקדמי עלילה'. לעומת זאת, לאירועים מסוימים מעמד מיוחד: השפעתם על העולם המיוצג בסיפור, ובמידה רבה גם על תכלית

הסיפור ועיקרו, רבה ועמוקה יותר; אלו הם, בלשוננו כאן, 'אירועים בעלי חשיבות מיוחדת' (בשפה מקצועית יותר נאמר כי אירועים אלה מממשים מידה גבוהה יותר של eventfulness). כאמור, איננו משתמשים כאן בקטגוריות אלה אלא כדוגמא בלבד, ואין לראותן אפילו כסיכום רופף של סוגיה תיאורטית סבוכה ומסועפת.⁹

11.21. בשלב זה ספריית התגים שלנו נראית כבתמונה שלהלן. שם הספרייה מופיע למעלה ('ניתוח עלילתי של סיפורים תלמודיים'); תחתיה מופיעות שתי מערכות תגים ('אירועים' ו'דמויות'), שעדיין אינן תגים בפני עצמן – ולכן נטולות צבע; ותחת כל מערכת מופיעים תגים מסוגים שונים (למשל: 'אירועים בעלי משמעות מיוחדת' או 'דמויות מספרים'), תג תג והצבע המיוחד לו. מתחת לאחד התגים ('דמויות פעילות בסיפור') מופיעים כמה תגי משנה ('דמות מעולמם של חז"ל', 'דמות מקראית', 'דמות אלהית', 'דמות אנונימית') המחולקים אף הם לצבעים שונים. תיאורטית, אפשר להמשיך לייצר עוד ועוד תגים, עוד ועוד תגי משנה, ואף תגי-משנה לתגי-משנה (כמעט ללא הגבלה). מערכת התגים הסופית עשויה להיות מסועפת מאוד, אך כאן, מטבע הדברים, נשאיר את הדוגמא בפשטותה.

ניתוח עלילתי של סיפורים תלמודיים	
Tagsets	Tag Type Color
אירועים	
אירועים בעלי משמעות מיוחדת	
אירועים מקדמי-עלילה	
דמויות	
דמויות מספרים	
דמויות פעילות בסיפור	
דמות מעולמם של חז"ל	
דמות מקראית	
דמות אלהית	
דמות אנונימית	

11.22. כדאי להתעכב על ההבדל העקרוני בין התגים השונים. יש תגים שאמורים להוסיף על גבי הטקסט מידע כמעט אובייקטיבי: כך למשל התגים שנועדו לקבוע אם גיבור הסיפור הוא 'דמות מקראית' או 'דמות אלהית'. לעומת זאת, יש תגים שנועדו לתעד תוצר של פרשנות סובייקטיבית, שאפשר בהחלט לחלוק עליה – למשל, התג 'אירועים בעלי משמעות מיוחדת'. לדידו של המחשב, כמובן, אין כל הבדל בין סוגים אלה: כולם נוספו על ידי המשתמש.

⁹ לסיכום מקיף ועדכני של הנושא ראו: Peter Hühn, ['Event and Eventfulness'](#), In: Hühn, Peter et al. (eds.): the living handbook of narratology. Hamburg: Hamburg University Press.

11.23. כמו כן, ארגון היחסים בין תגים לתגי-המשנה שלהם דורש אף הוא מחשבה: היכן עובר הגבול בין

קטגוריה לתת-קטגוריה, ולתת-קטגוריה? אם, מסיבה כלשהי, אנו צפויים להעדיף תמיד להשתמש בתגים הספציפיים יותר, איזו תועלת נצפה לקבל מהשארת קטגוריית-העל? האם עדיפה פריסה רוחבית מקבילה של התגים, או סידור הירארכי שלהם? התשובות לשאלות אלה משתנות ממחקר למחקר, ועשויות להשתנות גם במהלך המחקר עצמו; עצם העיסוק בהן בשלב הכנת ספריית התגים מעודד את החוקר למחשבה רפלקסיבית על מושגי היסוד המשמשים אותו ועל הזיקה ביניהם.

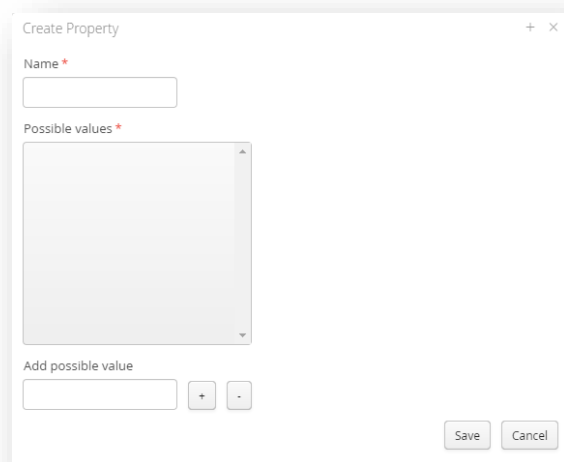
11.24. CATMA מציעה לחוקר אמצעי סיווג נוסף, שאינו מופיע בתצורה של תג-משנה, אלא כ'תכונה' (Property)

(Property) של תג (או תג-משנה) כלשהו. כדי להוסיף תכונה לתג נסמן את התג הרלוונטי, ונלחץ על כפתור

Create Property הנמצא בחלקו הימני-תחתון של מסך ה- Tag Type Manager.

11.25. בחלון החדש שנפתח נדרש להגדיר הן את המאפיין הכללי של התכונה, הן את ערכיה השונים. תחילה

נציין את שם התכונה בתיבת Name, ולאחר מכן נקליד את ערכיה השונים אחד אחרי השני בתיבת שתחת



Add possible value. לחיצה על סימן

החיבור ליד התיבת תוסיף את הערך שהצענו

לתיבת ה- Possible values. בגמר הוספת

ערכי התכונה נלחץ על Save.

11.26. בדוגמא שלפנינו נוכל להוסיף תכונה

לתג 'דמויות מספרים', אם, למשל, ברצוננו

לסווג את המספרים בהתאם למעמדם

בסיפור. לתכונה הכוללת

נקרא 'מעמד המספר

ביחס לעולם המיוצג',

ותחתיה נוסיף את

הערכים הבאים: 'מספר

כל יודע' – כלומר, מספר

היודע את כל אשר נעשה

דמויות	
דמויות מספרים	
מעמד המספר ביחס לעולם המיוצג	
דמויות פעילות בסיפור	
דמות אלוהית	
דמות אנונימית	
דמות מעולמם של חז"ל	
דמות מקראית	

בסיפור, בכל מקום ובכל זמן, ויודע אף את אשר בלב ובדעת גיבוריו – ו'מספר מעורב', כלומר, מספר הממלא תפקיד בעולם המיוצג בסיפור, ובדרך כלל אינו יודע דבר מעבר לנקודת מבטו המוגבלת (לעתים מבחינים בין 'מספר עֵד' ל'מספר גיבור', אך כאן נפשט את הדיון ונסתפק בהגדרות שטחיות בלבד).

11.27. לאחר לחיצה על Save תצטרף התכונה לספריית התגים במקום המתאים, ותופיע אך ורק בשמה הכולל. ערכיה השונים יוצגו בפנינו שוב כשנתייג קטע מסוים בטקסט כ'דמויות מספרים', ונתבקש לקבוע איזו תכונה של התג מגולמת בו. יושם אל לב שבתחום התכונות איננו מגדירים צבעים: האפשרות לסווג קטע מסוים בטקסט כבעל תכונה מסוימת תעמוד בפנינו רק כאשר נסמן את אותו קטע בתג הרלוונטי (בשונה מתגי-המשנה, שלהם צבעים עצמאיים). מבחינה מעשית חשוב לזכור, אם כן, כי כאשר נתבונן לאחר מעשה בטקסט המתוייג, נראה תגים בלבד – בעוד התכונות תיוותרנה חבויות ברקע. מלבד זאת, בניגוד למערך המשלב תגים ותגי-משנה, תכונות אינן מאפשרות סיווג משני נוסף.

11.28. מכל בחינה אחרת, אין כמעט הבדל מהותי בין תגי המשנה לתכונות. אלו הן שתי צורות לסיווג מתקדם של תופעות טקסטואליות. הבחירה בכל אחת מהן היא, בעיקרה, עניין של סגנון והעדפה אישית, הנרכשת לאחר התנסות ממושכת במערכת. היא עשויה כמובן גם להיות מושפעת מאופיו של המחקר.

11.29. כל אחד ממרכיבי ספריית התגים – מערכות התגים, סוגי התגים, תגי המשנה וסימני התכונות – יכולים להיערך מחדש או להימחק בכל רגע נתון, בהתאם לרצון החוקר, בעזרת כפתורי ה- Remove וה- Edit הנמצאים בצידו הימני של מסך ה- Tag Type Manager.

11.30. חשוב מכל: יושם אל לב עד כמה היינו חופשיים לגמרי לעצב קטגוריות פרשניות כרצוננו; במקרה זה גיבשנו ספריית תגים המשלבת פרספקטיבה נראטולוגית ומותאמת לחקר הסיפור התלמודי, אך באותה מידה ממש אפשר היה גם לגבש ספריה שונה לגמרי, בכל תחום עיוני כמעט העולה על הדעת, ובלבד שיש לו זיקה כלשהי לטקסטים.

11.31. כפי שנראה בהמשך, הגדרת מערכת תגים אינה בבחינת 'סוף פסוק'. אין בה כדי לקבע מהלך דינאמי שעשוי ללבוש צורה ולפשוט צורה בכל שלבי העבודה. כך, למשל, חוקרים רבים ימצאו תועלת בתיוג ראשוני שטחי, שאינו מגיע כלל לרזולוציה ספציפית-מדי בדמותם של תגי-משנה ותכונות. אין זה רעיון רע כל כך: הוא מאפשר גמישות רבה יותר, ואף מקל על המשתמש לחזור בהמשך לסבב תיוג נוסף, מדויק ומפורט יותר.

12. הכנת מסך התיוג (Tagger): טעינת טקסט, מערכת תגים ואסופת תיוג

12.1. עם השלמת בניית ספריית התגים – השלמה חלקית וזמנית בלבד, כאמור לעיל, משום שבכל רגע נוכל לשוב

ולשכלל אותה – נגיש להכין את הטקסט לתיוג. בפינה הימנית-עליונה של מסך ה- Tag Type Manager

ממוקם כפתור שתפקידו 'לטעון' את מערכת התגים אל הטקסט המטופל כעת: Load Tagset into currently

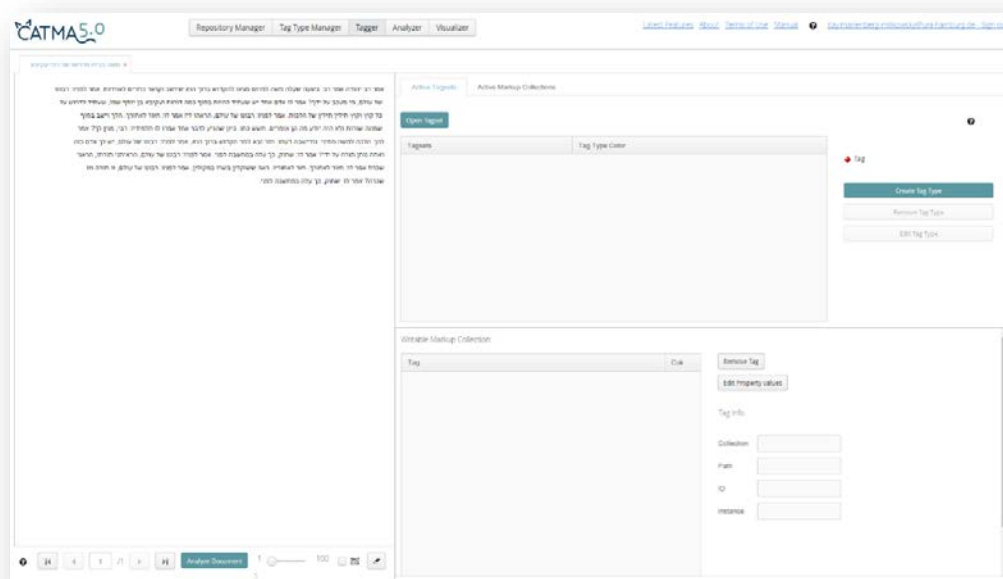
active Document. הואיל ועדיין לא פתחנו טקסט – כלומר, עדיין לא הפכנו את אחד הטקסטים הזמינים

במערכת לטקסט פעיל, שאמור להיות נדון במסך התיוג – אפשרות זו חסומה בפנינו.

12.2. כדי לפתוח טקסט במערכת, נחזור למסך הראשי, מסך ה- Repository Manager. במרכז המסך ממוקמת,

כזכור, רשימת המסמכים. נסמן את הטקסט המבוקש, ונלחץ על כפתור Open Document מתחת

לרשימה. פעולה זו תוביל אותנו לראשונה למסך ה- Tagger.



12.3. בצידו השמאלי של המסך נמצא הטקסט (בדוגמא שבתמונה: סיפור משה בבית מדרשו של רבי עקיבא). טקסט

ארוך יותר יחולק למקטעים שווים באורכם (עד 100 מקטעים לכל היותר), שהגישה אליהם אפשרית בעזרת

כפתורי הדפדוף שמתחת לטקסט. החוקר יכול לשלוט במספר המקטעים השווים בעזרת הזזת העיגול הקטן

הנמצא על הסקאלה שבין 1 ל- 100, הנתונה אף היא מתחת לטקסט. במקרים מסוימים עשויים טקסטים ארוכים

להופיע עם פס גלילה אנכי, ויש לשים לב לכך בזמן התיוג על מנת לא להחמיץ חלקים נסתרים.

12.4. בצידו הימני של המסך נמצא את המנגנונים השונים הנדרשים לתיוג הטקסט. המשתמש המתחיל עשוי לזהות

כעת מדוע מנגנונים אלה נראים ריקים, ומדוע אין הוא יכול לתייג – אף על פי שספריית התגים כבר גובשה מזמן;

מערכת תגים | Tag-set

אך כאן אנו פוגשים שוב את עקרון ההפרדה בין השכבות

אסופת תיוג | Markup Collection

השונות ב-CATMA: אמנם העלנו את הטקסט למסך ה-

טקסט | Text



Tagger, אך עדיין לא העלנו למסך זה את מערכות התגים.

12.5. בבואנו לעשות זאת עומדות בפנינו שתי אפשרויות: (א) נוכל לשוב למסך ה-Tag Type Manager, לסמן את

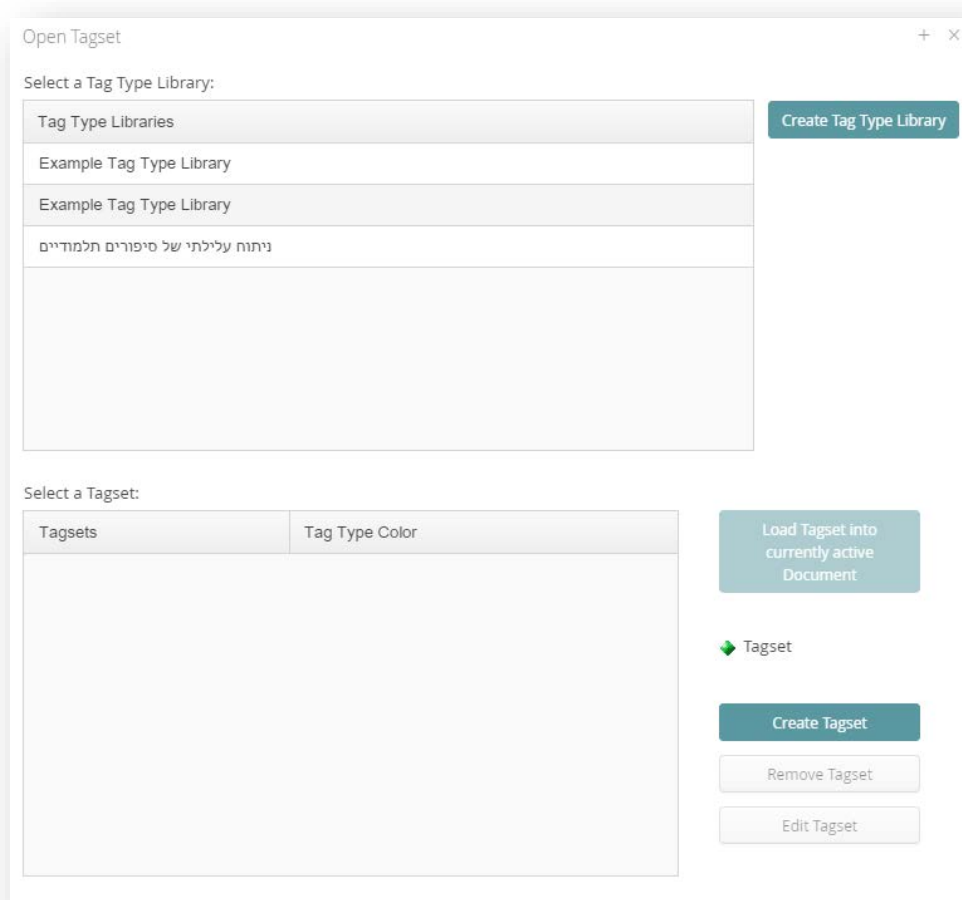
מערכת התגים המבוקשת, וללחוץ על כפתור Load Tagset into currently active Document הנמצא

בפינה הימנית-עליונה של המסך (אפשרות זו נמנעה מאיתנו כל עוד לא טענו טקסט למסך ה-Tagger; אך

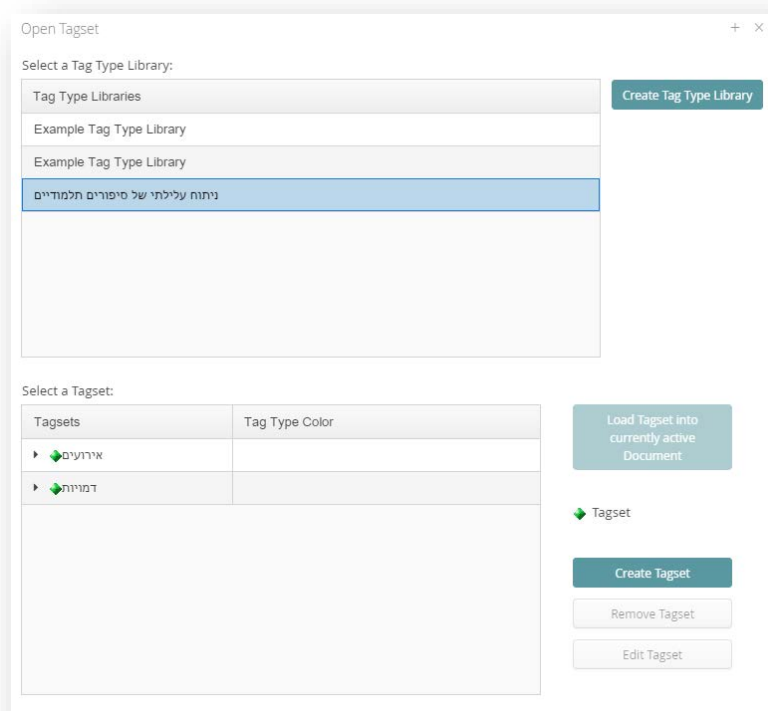
משהטקסט נטען, שוב איננו מנועים מלעשות זאת). לאחר מכן נשוב למסך ה-Tagger, ונמצא את מערכות

התגים בצידו הימני. (ב) נוכל להישאר במסך ה-Tagger ולטעון את מערכות התגים ישירות מתוכו. לשם כך

נלחץ על כפתור Open Tagset הנמצא במרכז המסך בחלקו העליון, ונקבל את החלון הבא:



12.6. יוזכר כי לפנינו מערכת הירארכית בעלת הגיון פנימי משלה. לפיכך נסמן תחילה את ספריית התגים הרלוונטית (Tag Type Library). בדוגמא זו: 'ניתוח עלילתי של סיפורים תלמודיים'. כעת מופיעות למטה מערכות התגים (Tagsets) שיצרנו קודם:



12.7. נסמן את אחת המערכות

(בדוגמא זו: 'אירועים')

או 'דמויות', ונלחץ על

Load Tagset

into currently

.active Document

לחיצה זו מחזירה אותנו

אל מסך ה- Tagger,

וכעת במרכזו, בחלקו

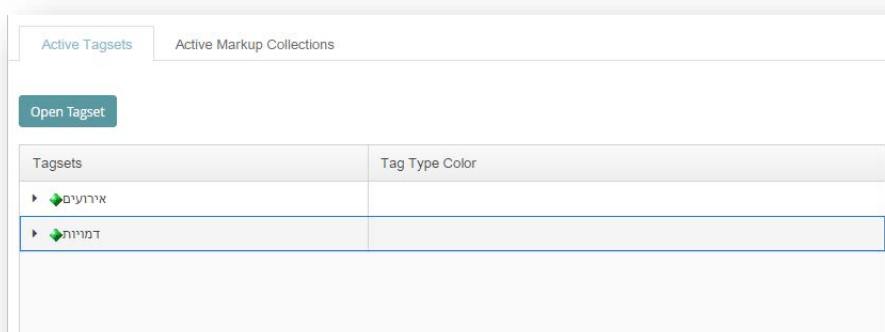
העליון, תופיע מערכת

התגים שטענו כעת.

נחזור על התהליך כדי

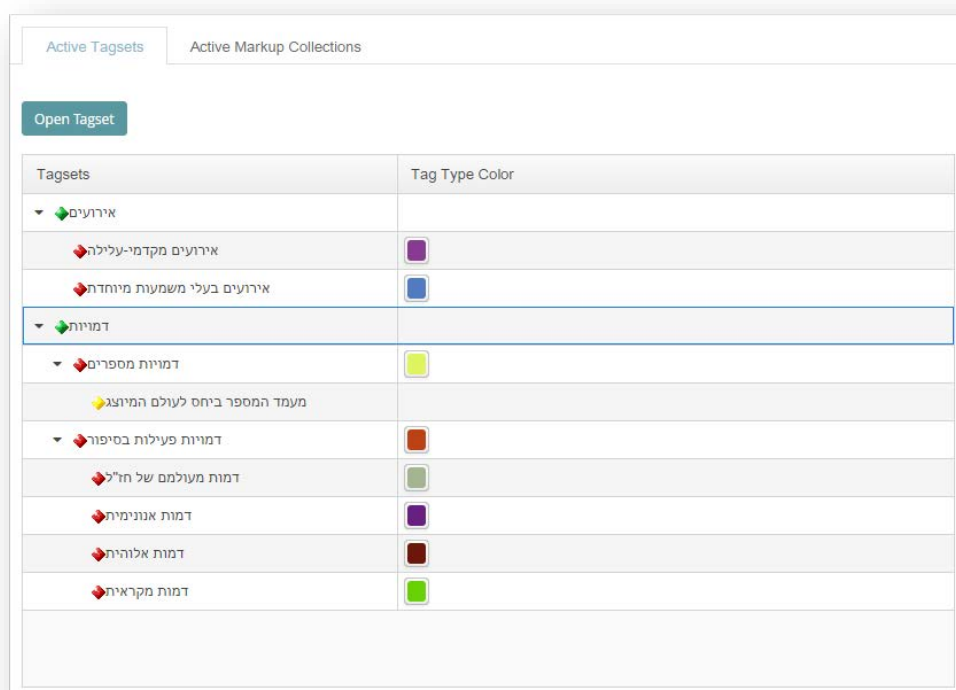
לסמן גם את המערכת השנייה.

12.8. בשלב זה מופיעות שתי מערכות התגים שטענו במרכז מסך ה- Tagger, תחת הכותרת Active Tagsets:



12.9. לחיצה על המשולשים הקטנים שלצידן תציג בפנינו את המערכות במלואן, על שלל צבעיהן – כפי שהגדרנו,

ארגנו ועיצבנו אותן קודם לכן במסך ה- Tag Type Manager.



12.10. צבעי התגים נראים הפעם ככפתורים קטנים, ומזמינים אותנו להשתמש בהם לתיג הטקסט. אולם חוקר

שינסה לתייג כעת, לא יצליח. בנקודה זו המשתמש-המתחיל הממוצע תוהה, לא בלי תסכול, מדוע עדיין אינו

יכול לתייג: הוא הרי טען כבר טקסט, הוא טען גם מערכות תיג – מה עוד חסר?

12.11. שוב אנו ניצבים, כצפוי, מול עקרון הפרדת השכבות של CATMA. אנו זקוקים ל'מקום' וירטואלי שבו

מערכת תגים Tag-set



נשמור את שכבת התיג שלנו על גבי הטקסט. לשם כך עלינו

ליצור אסופת תיג. לצד הלשונית Active Tagsets ממוקמת

אסופת תיג | Markup Collection

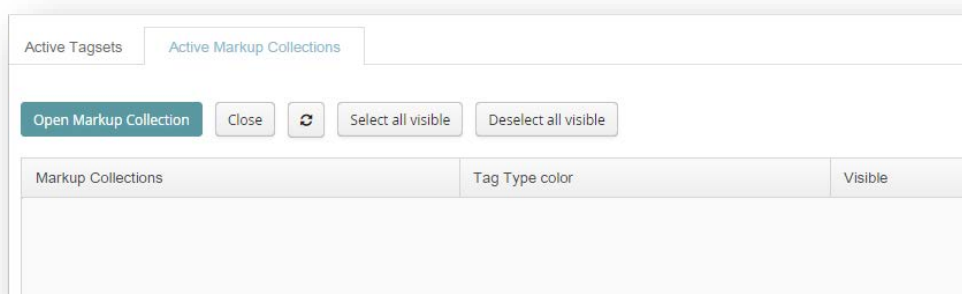
לשונית נוספת: Active Markup Collections. לחיצה

טקסט | Text

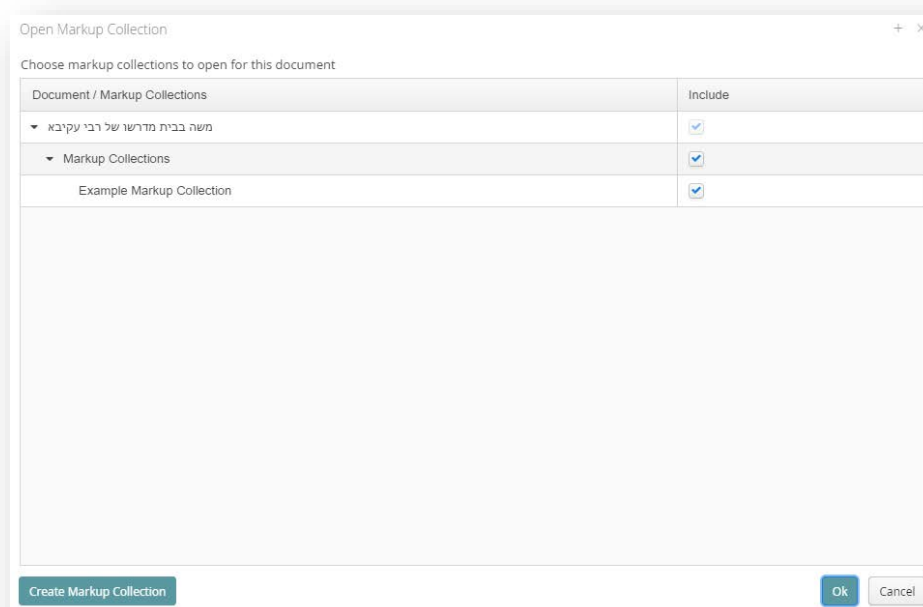


עליה תוביל אותנו למקום שבו נוכל לנהל את אסופות התיג

השונות.



12.12. לחיצה על כפתור Open Markup Collection תוביל אותנו אל החלון הבא:



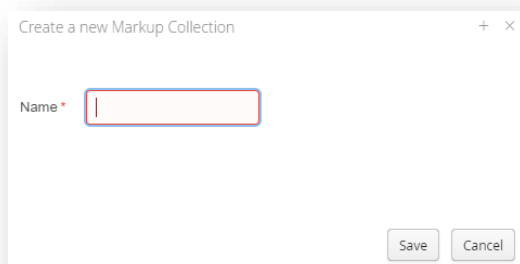
12.13. ראשית יושם אל לב, כי אסופות תיוג כפופות לטקסט מסוים אחד: נקודת המוצא במבנה ההירארכי

שלפנינו היא הטקסט עצמו (במקרה זה: 'משה בבית מדרשו של רבי עקיבא'), וממנה יכולות להסתעף אסופות-

תיוג רבות, תחת הכותרת הכללית Markup Collections. צורת הרבים כאן אינה מקרית: המשתמש יכול להזין

כמה אסופות תיוג שירצה לטיפול בטקסט מסוים אחד, אם לצרכי אותו מחקר, ואם לצרכי מחקר אחר.

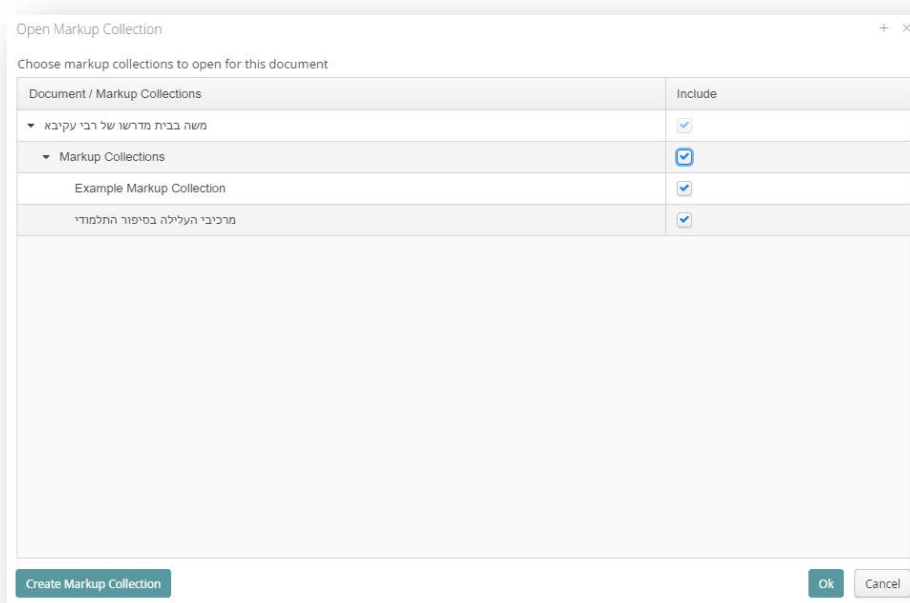
12.14. בשלב הנוכחי אנו זקוקים אפוא לאסופת תיוג יעודית אחת, ולכן נלחץ על כפתור Create Markup



Collection שבתחתית חלון זה. שוב ייפתח

בפנינו חלון קביעת שם – והפעם, קביעת שם לאסופת התיוג שבה נשמור את שכבת התיוג המסוימת של הטקסט המסויים שעליו אנו עובדים כעת. בהתחשב בדוגמא המעסיקה אותנו כאן, נכנה את אסופת התיוג 'מרכיבי העלילה בסיפור התלמודי', ולבסוף נלחץ על Save.

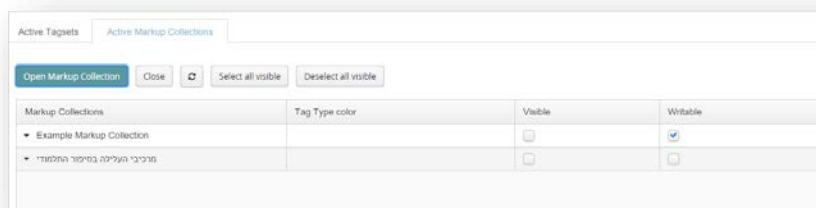
12.15. לחיצה זו מחזירה אותנו אל החלון שבו היינו קודם – אך הפעם, האסופה שהגדרנו זה עתה נוספה לרשימה.



12.16. נוודא שהאסופה המיועדת מסומנת ב-√ (נתעלם לעת עתה מתפקידן של האסופות האחרות), ונלחץ על

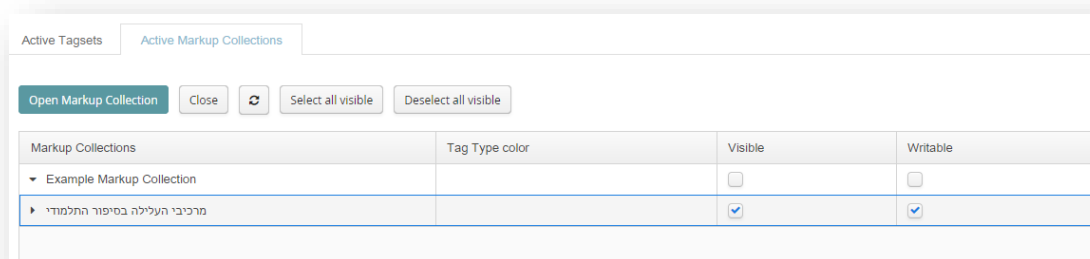
כפתור ok. לחיצה זו מחזירה אותנו לחלון ה-Tagger, וכעת אסופת התיוג שהגדרנו מופיעה במרכז, תחת

הכותרת Markup Collections.



12.17. אנחנו כבר קרובים מאוד לתחילת התיוג; כל שנותר לעשות הוא לסמן ✓ בתיבות שמשמאל לאסופה

שהגדרנו – הן בתיבת ה-Visible, המנחה את המערכת להראות את כל התיוג שנוצר עד כה (אפשרות זו חשובה בעיקר כשחוזרים וטוענים טקסט שכבר תויג. בשלב הראשוני המתואר כאן אין לכך כמעט חשיבות), והן בתיבת ה-Writable, המנחה את המערכת לאפשר לנו ליצור תגים הלכה למעשה באסופת התיוג שהוגדרה.



12.18. כעת נחזור ללשונית ה-Active Tagsets, על מנת לראות את מערכות התגים שבהן נשתמש מיד. **המסך**

שלפנינו ילווה אותנו לאורך רוב תהליך התיוג: בצידו השמאלי, הטקסט שנטען למערכת; בצידו הימני, מערכות התיוג שנטענו (וסביבן כלי עבודה שונים לניהול התיוג); וברקע, אסופת התיוג שבה יישמרו תיוגינו. לחיצה עם הסמן על קוי ההפרדה בין האזורים השונים במסך תאפשר לנו להזיז אותם ולשנות מעט את החלוקה הפנימית של המסך (אפשרות זו מועילה במיוחד כאשר עובדים עם מערכת תיוג גדולה ומסועפת, ומבקשים לראות כמה שיותר ממנה בבת אחת).

12.19. שלוש השכבות פעילות אפוא במסך ה-Tagger; והנה, בשעה טובה ומוצלחת, אנו יכולים סופסוף

להתחיל לתייג. לאמיתו של דבר, סדרת הפעולות שתוארה בפרק זה באריכות כלשהי מתבצעת במהירות רבה,

וככל שהמחקר מתקדם ומסתעף כך נהנים מיתרונות אופיה המודולארי. אולם בשלבי ההתנסות הראשוניים כדאי לחזור ולתרגל אותה, משום שהקפדה על ביצוע שלם שלה, לפרטי פרטיה, היא תנאי הכרחי למימוש הפוטנציאל של המערכת.

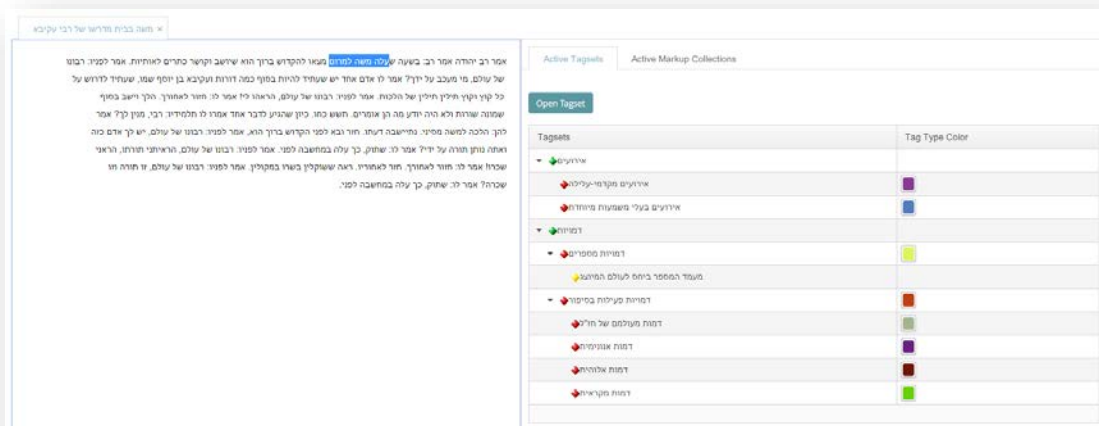
מערכת תגים Tag-set ✓

אסופת תיוג Markup Collection | ✓

טקסט Text | ✓

13.1. עתה, משמסך ה- Tagger מוכן לעבודה, נוכל להתחיל לתייג. מבחינה טכנית, פעולת התיוג פשוטה להפליא:

עלינו לסמן מקטע מבוקש בטקסט וללחוץ על צבע התג הרלוונטי. התג המבוקש יתווסף תחת הטקסט.



13.2. כך למשל בדוגמא שלפנינו, אם ברצוננו להצמיד למילים 'עלה משה למרום' את התג 'אירועים מקדמי-עלילה',

עלינו לסמן את המילים המבוקשות (בדיוק כפי שמסמנים קטעי טקסט בכל תוכנה אחרת) וללחוץ על קוביית

הצבע הסגול הנמצאת מימין לתג 'אירועים מקדמי-עלילה'. הצבע הסגול יתווסף תחת הטקסט במקום המתאים

(הפס האדום הדק שתחת התג יעלם כשנעבור לתג הבא. כל תכליתו היא לסמן באיזה תג אנו עוסקים כעת).

אמר רב יהודה אמר רב: בשעה שעלה משה למרום מצאו להקדוש ברוך הוא שיושב וקושר כתרים לאותיות. אמר לפניו: רבנו של עולם, מי מעבך על ידיך? אמר לו אדם אחד יש שעתידי להיות בסוף כמה דורות ועקיבא בן יוסף שמו, שעתידי לדרוש על כל קוץ וקוץ תילין תילין של הלכות. אמר לפניו: רבנו של עולם, הראהו לי! אמר לו: חזור לאחורך. הלך וישב בסוף שמונה שורות ולא היה יודע מה הן אומרים. תשש כחו. כיון שהגיע לדבר אחד אמרו לו תלמידיו: רבי, מנין לך? אמר להן: הלכה למשה מסיני. נתישבה דעתו. חזר ובא לפני הקדוש ברוך הוא, אמר לפניו: רבנו של עולם, יש לך אדם כזה ואתה נותן תורה על ידי? אמר לו: שתוק, כך עלה במחשבה לפני. אמר לפניו: רבנו של עולם, הראיתני תורתו, הראני שכרה! אמר לו: חזור לאחורך. חזר לאחוריו. ראה ששוקלין בשרו במקלין. אמר לפניו: רבנו של עולם, זו תורה חזו שכרה? אמר לו: שתוק, כך עלה במחשבה לפני.

13.3. בדרך זו נמשיך ונתייג את הטקסט כולו. כך למשל נראה הטקסט המתוייג לאחר שהוספנו לו את התגים 'אירועים

מקדמי-עלילה' (סגול) ו'אירועים בעלי משמעות מיוחדת' (כחול):

אמר רב יהודה אמר רב: בשעה שעלה משה למרום מצאו להקדוש ברוך הוא שיושב וקושר כתרים לאותיות. אמר לפניו: רבוננו של עולם, מי מעכב על ידיך? אמר לו אדם אחד יש שעתידי להיות בסוף כמה דורות ועקיבא בן יוסף שמו, שעתידי לדרוש על כל קוץ וקוץ תילין תילין של הלכות. אמר לפניו: רבוננו של עולם, הראהו לי! אמר לו: חזור לאחורך. הלך וישב בסוף שמונה שורות ולא היה יודע מה הן אומרים. תשש כחו. כיון שהגיע לדבר אחד אמרו לו תלמידיו: רבי, מנין לך? אמר להן: הלכה למשה מסיני. נתיישבה דעתו. חזר ובא לפני הקדוש ברוך הוא, אמר לפניו: רבוננו של עולם, יש לך אדם כזה ואתה נותן תורה על ידי? אמר לו: שתוק, כך עלה במחשבה לפני. אמר לפניו: רבוננו של עולם, הראיתני תורתו, הראני שכרה! אמר לו: חזור לאחורך. חזר לאחוריו. ראה ששוקלין בשרו במקלין. אמר לפניו: רבוננו של עולם, זו תורה חו שכרה? אמר לו: שתוק, כך עלה במחשבה לפני.

13.4. לא כאן המקום לדון בהכרעות הפרשניות המגולמות בתיוג זה, וברור שאפשר גם לחלוק עליהן. לצורך ענייננו

כעת, די אם נאמר שפרשנות זו מניחה כי החוויה המטלטלת שמשה עובר בבית מדרשו של רבי עקיבא, שם "לא

היה יודע מה הן אומרים" עד ש"תשש כחו", ורק לאחר ששמע כי מה שנאמר מיוחס לו עצמו "נתיישבה דעתו",

נראים רבי-משמעות יותר מהעליה למרום לבדה, או מתנועת הרצוא-ושוב של משה בין שמים וארץ, בין עבר

ועתיד, אליה המספר מתייחס כמעט כבדרך אגב. הכרעה פרשנית נוספת היא לראות בכל אמירה, בכל דיאלוג,


אירוע, אם גם מינורי; בהמשך נשוב לעסוק בה כדי להציג שיטת תיוג נוספת.

13.5. לאחר שנתייג גם את שמות הדמויות בתגים המתאימים ('דמות מעולמם של חז"ל', 'דמות אנונימית', 'דמות

מקראית', 'דמות אלוהית'), נראה הטקסט המתוייג כך:

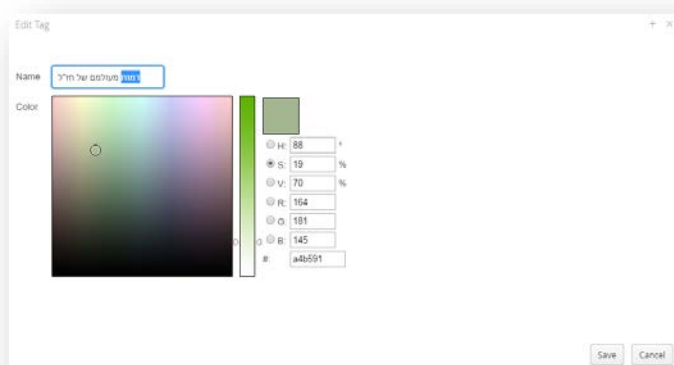
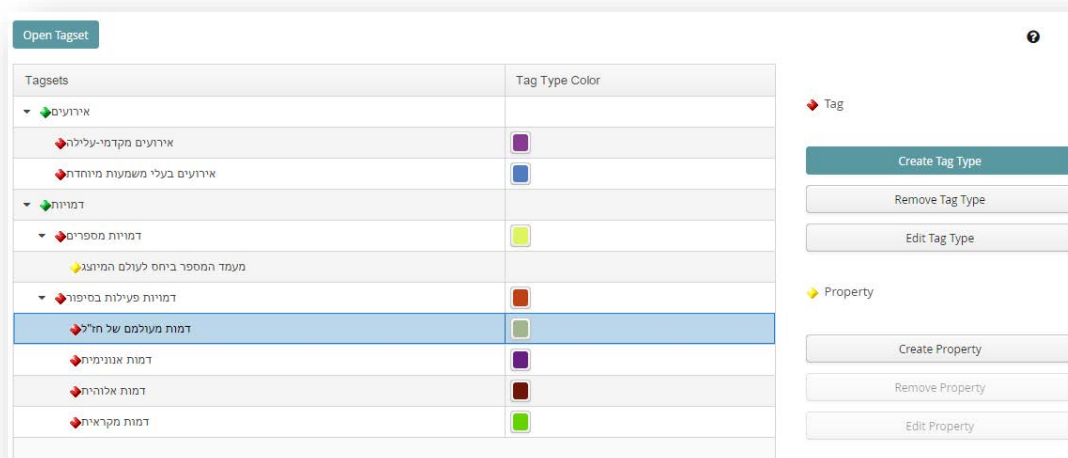
אמר רב יהודה אמר רב: בשעה שעלה משה למרום מצאו להקדוש ברוך הוא שיושב וקושר כתרים לאותיות. אמר לפניו: רבוננו של עולם, מי מעכב על ידיך? אמר לו אדם אחד יש שעתידי להיות בסוף כמה דורות ועקיבא בן יוסף שמו, שעתידי לדרוש על כל קוץ וקוץ תילין תילין של הלכות. אמר לפניו: רבוננו של עולם, הראהו לי! אמר לו: חזור לאחורך. הלך וישב בסוף שמונה שורות ולא היה יודע מה הן אומרים. תשש כחו. כיון שהגיע לדבר אחד אמרו לו תלמידיו: רבי, מנין לך? אמר להן: הלכה למשה מסיני. נתיישבה דעתו. חזר ובא לפני הקדוש ברוך הוא, אמר לפניו: רבוננו של עולם, יש לך אדם כזה ואתה נותן תורה על ידי? אמר לו: שתוק, כך עלה במחשבה לפני. אמר לפניו: רבוננו של עולם, הראיתני תורתו, הראני שכרה! אמר לו: חזור לאחורך. חזר לאחוריו. ראה ששוקלין בשרו במקלין. אמר לפניו: רבוננו של עולם, זו תורה חו שכרה? אמר לו: שתוק, כך עלה במחשבה לפני.

13.6. עקרון הרב-שכבתיות של CATMA מודגם כאן היטב: שתי מערכות תגים מונחות זו על גב זו, ללא כל תלות הכרחית אחת בשניה. בהמשך נלמד כיצד לנצל עקרון זה לצרכינו. אולם גם תמונה זו לבדה מעוררת שאלות קונספטואליות. מה בדיוק כדאי לתייג? האם נסמן 'דמויות' רק בהתייחס למופעים של שמותיהן, כפי שמוצע כאן? אולי עדיף לסמן באותו תג את כל מעשייה (ודבריה) של הדמות? משה הוא דוגמא טובה לכך: שמו כמעט לא מוזכר בסיפור, אף על פי שברור כי הוא הגיבור המרכזי. לשאלות אלה אין הכרעה ברורה משום שהן תלויות במגמת המחקר וביעדיו; אך מוטב לחשוב עליהן בכל שלב של המחקר, ובמידת הצורך לעדכן את מערכות התגים ואת שיטת התיוג.

13.7. יוער גם כי CATMA מאפשרת תיוג לא-רציף. אם ברצוננו לתייג מילים שאינן סמוכות זו לזו בתג אחד – כלומר, לא רק כשני ביטויים לתופעה אחת, אלא כתופעה אחת המופיעה במקוטע בגוף הטקסט – נסמן $\sqrt{\quad}$ בריבוע הקטן הממוקם תחת צידו הימני של הטקסט (allow multiple discontinuous selections).  לאחר מכן נוכל לסמן מקטע בטקסט, ואז מקטע נוסף (וכן הלאה, בהתאם לצורך); רק לאחר לחיצה על אחד מכפתורי התגים, יקבלו מקטעים אלה את התיוג הנבחר, תוך דילוג על מקטעי הטקסט החוצצים ביניהם. כאשר נרצה לשוב ולתייג ברציפות נסיר את סימן ה- $\sqrt{\quad}$.

13.8. לעתים מתעורר ספק; מחשבה מחודשת נדרשת לבירור הגדרות היסוד עימן ניגשנו אל הטקסט. כך למשל 'עקיבא בן יוסף' ו'תלמידיו': את שניהם סימנו לעיל בתג 'דמות מעולמם של חז"ל', אך אפשר לדייק יותר: הראשון הוא חכם (וליתר דיוק: תנא) ידוע, בעוד השניים הם תלמידים אנונימיים. בה בעת, התג 'דמות אנונימית' שהוכן מראש, אינו מספק – שכן הללו לא לגמרי אנונימיים, אלא תלמידי רבי עקיבא. להלן נדגים כיצד אפשר לשפר את התיוג.

13.9. אם ברצוננו לשנות את הגדרות היסוד של תג מסוים, נפנה לצידו הימני של המסך – אל המקום שבו מונחות לפנינו מערכות התגים הפעילות. נסמן את התג המבוקש, ונלחץ על כפתור Edit Tag Type.



13.10. בחלון שנפתח נוכל

לשנות הן את שם התג, והן את

צבעו. לאחר מכן נלחץ על Save.

בדוגמא שלפנינו נותיר את הצבע

כמות שהוא, אך את שם התג נשנה

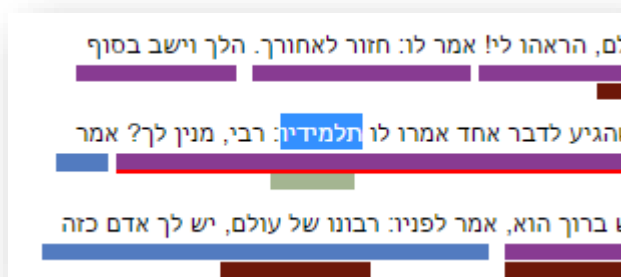
מ'דמות מעולמם של חז"ל'

ל'חכם מעולמם של חז"ל'.

13.11. במידת הצורך, אם נרצה למחוק תג לגמרי – כלומר, למחוק אותו מרשימת התגים במערכת התגים –

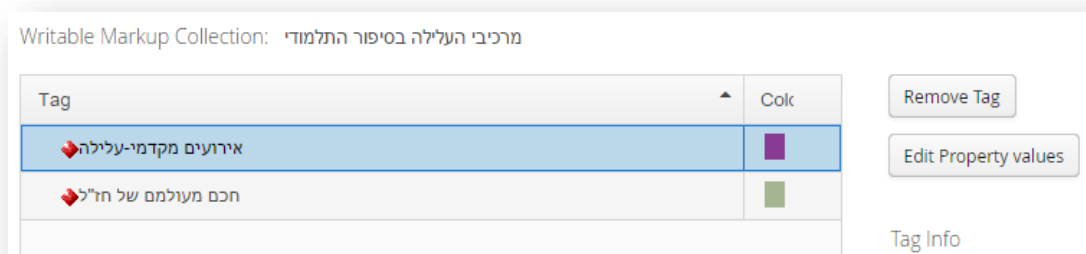
נלחץ על כפתור Remove Tag Type.

13.12. אולם רצוננו, בדוגמא הנוכחית, איננו למחוק תג ממערכת התגים, כי אם למחוק תיוג מאסופת התיוג,



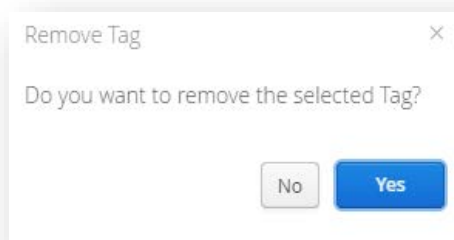
כלומר – להסיר את התג 'דמות מעולמם של חז"ל' (שהפך בינתיים ל'חכם מעולמם של חז"ל') מהמילה 'תלמידיו'. לשם כך נחזור אל הטקסט בצידו השמאלי של המסך, ונסמן את המילה המבוקשת, ממש כאילו ביקשנו לתייגה מחדש – אך הפעם, לקראת מחיקת התג.

13.13. מחיקת תיוגים מתנהלת באזור הימני-תחתון של המסך:



13.14. הואיל ומקטע הטקסט שסומן נושא על גביו, לפי שעה, שני תגים שונים, שניהם מופיעים ברשימה

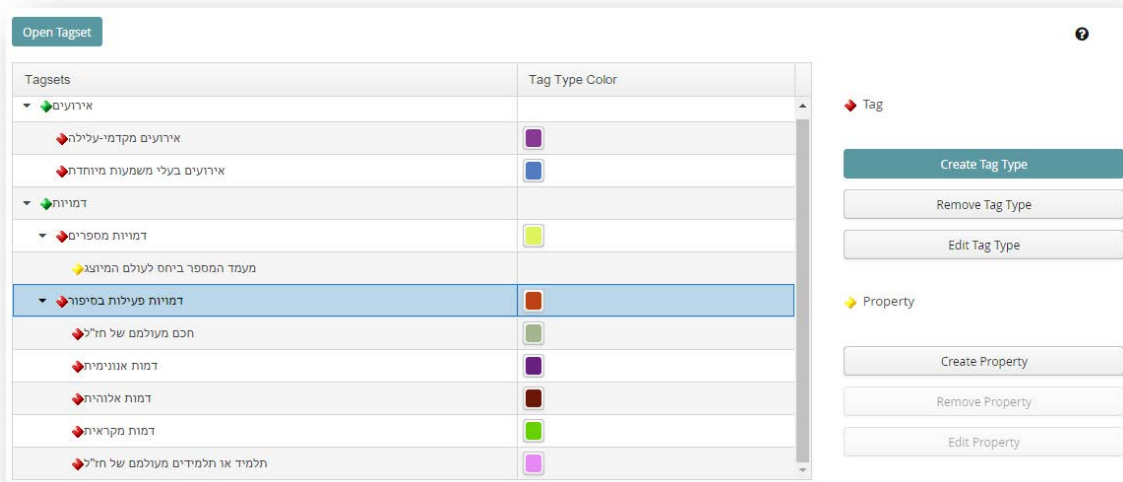
הנפתחת באזור זה של המסך. נסמן את התג שאותו ברצוננו למחוק – במקרה זה, התג 'חכם מעולמם של חז"ל', ונלחץ על Remove Tag (יושם אל לב: ברירת המחדל של המערכת היא לסמן את התג הראשון ברשימה –



כפי שנראה בתמונת המסך לעיל – ולכן חשוב להקפיד לבדוק מהו התג המסומן לפני הלחיצה על Remove Tag). לחיצה זו תוביל אותנו לתיבת אזהרה – האם באמת אתם מעוניינים למחוק את התג המסומן? – ולחיצה על Yes תאשר סופית את המחיקה. פעולה זו עשויה להיארך כמה שניות.

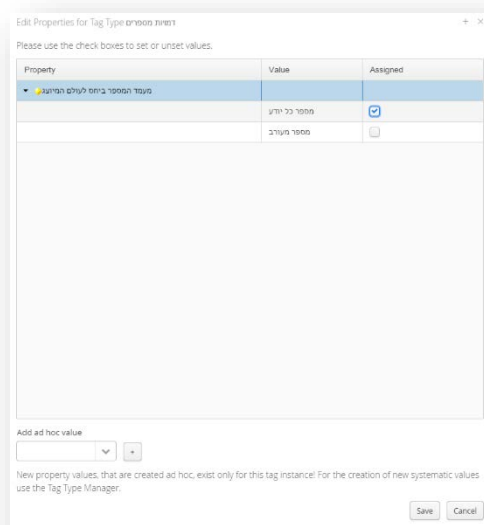
13.15. כעת נבקש לייצר תג חלופי: 'תלמיד או תלמידים מעולמם של חז"ל'. אנו מבקשים להעמיד תג זה כתג-

משנה לתג 'דמויות פעילות בסיפור'; לפיכך נסמן את תג-העל 'דמויות פעילות בסיפור', ונלחץ על כפתור Create Tag Type הממוקם בצידו הימני של המסך. בחלון החדש שנפתח נגדיר את שם התג החדש ואת צבעו, ממש כפי שעשינו במסך ה- Tag Type Manager. התג החדש נוסף למערכת התיוג כולה – וימשיך להופיע בה כל אימת שנשתמש בה, ללא כל תלות בטקסט המסויים שלפנינו.



13.16. כדי להצמיד תג זה לטקסט, נסמן שוב את המילה המבוקשת בטקסט (במקרה שלנו, 'תלמיד'), ונלחץ על כפתור הצבע החדש שנוצר ליד התג החדש (במקרה שלנו, כפתור ורוד).

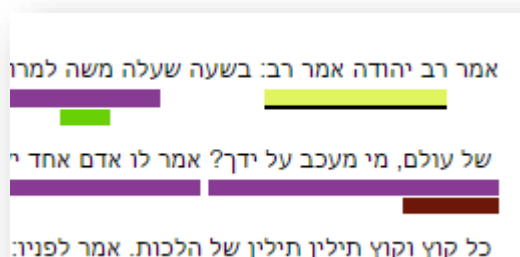
13.17. כזכור, מערכת התגים שבנינו כוללת לא רק תגים ותגי-משנה, אלא גם 'תכונות' (Properties). כך למשל הוספנו לתג 'דמויות מספרים' את התכונה 'מעמד המספר ביחס לעולם המיוצג', הנושאת בתוכה את הערכים



הבאים: 'מספר כל יודע' או 'מספר מעורב'. ואכן, אם נסמן את שמו של המספר – 'רב יהודה אמר רב' (כלומר, רב יהודה מוסר את הסיפור בשם רב) – בתג 'דמויות מספרים', יוצג מיד חלון קביעת התכונות:

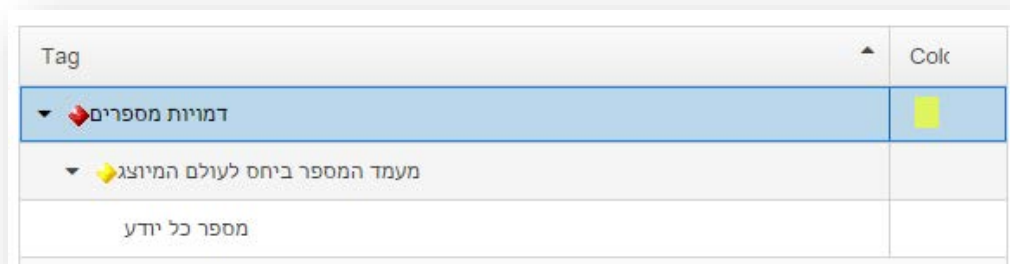
13.18. נסמן כעת את הערך הרצוי (במקרה זה, 'מספר כל יודע', משום שרב יהודה בשם רב אינו משתתף פעיל בעלילה, אלא מי שמספר עליה מבחוץ, ויודע כל שמתרחש – הן במהלכים הפיזיים

של הדמויות, הן במהלכים הנפשיים שלהן). לאחר שנלחץ על Save, התג יתווסף למילים המבוקשות, כאשר הוא נושא בחובו את התכונה והערך שייחסנו לו. עם זאת, בניגוד לתגי-משנה, אין התכונות מופיעות באופן ויזואלי על גבי הטקסט המתויג, אלא רק נשמרות בזכרון המערכת. אפשר לראות אותן בחלקו הימני-תחתון של המסך, וכמובן גם לנתחן לאחר מכן.

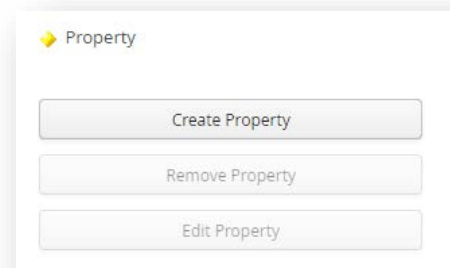


13.19. במידת הצורך נוכל אף להוסיף ערכים לתכונה נתונה במהלך פעולת התיוג, אם הערכים הכלולים בה כבר אינם מספקים. לשם כך נקליד את את הערך המבוקש בתיבה 'Add ad hoc value', ונלחץ על סימן הפלוס שלידו. עם זאת,

כפי שמוכן משמו, ערך זה יתווסף אד-הוק באופן מקומי בלבד. כדי לשנות תכונה באופן סיסטמטי יותר יש לטפל בה במסך ה- Tag Type Manager, כפי שהוסבר לעיל.



13.20. מבחינה עקרונית, כל האפשרויות שהודגמו קודם ביחס לתגים ולתגי-המשנה – עריכה, מחיקה והוספה – מוצעות גם ביחס לתכונות. אפשר לערוך תכונה, למחוק אותה ולהוסיף אותה בכל שלב שנבקש, בעזרת הכפתורים הרלוונטיים בצידו הימני של המסך.



13.21. נניח לפי שעה את הטקסט הראשון, ונתרגל את אפשרויות התיוג השונות בטקסט השני. לשם כך נחזור למסך ה- Repository Manager, נסמן את שם הטקסט המבוקש ברשימת הטקסטים שבמרכז המסך, ונלחץ על כפתור Open Document. שני הטקסטים הפתוחים יופיעו כעת תחת לשוניות נפרדות במסך ה- Tagger. לחיצה על הלשוניות מאפשרת מעבר בין הטקסטים השונים.

× משה בבית מדרשו של רבי עקיבא

× משה לא אומר שלום לקב"ה

ואמר רבי יהושע בן לוי: בשעה שעלה משה למרום מצאו להקדוש ברוך הוא שהיה קושר כתרים לאותיות. אמר לו: משה, אין שלום בעירך? אמר לפניו: כלום יש עבד שנותן שלום לרבו? אמר לו: היה לך לעזרני. מיד אמר לו: 'ועתה יגדל נא כח ה' כאשר דברת'.

13.22. יושם אל לב: גם אם אנו רואים טקסטים שונים כצדדים שונים של פרוייקט אחד, CATMA תופסת

אותם כקורפוס אך ורק בכל הנוגע למגוון הפעולות האפשריות במסך ה-Analyzer. בשלב התיוג במסך ה-Tagger, לעומת זאת, כל טקסט מסווג בנפרד, עם אסופת תיוג משלו. לפיכך עלינו להגדיר מחדש אסופת תיוג (לשם הנוחות, אפשר לכנות אותה בשם זהה לזו של הטקסט הראשון), ולטעון מערכות תגים לפעולה.

13.23. שימוש במערכת תגים אחת ליותר מטקסט אחד עשוי להעלות שאלות חדשות לדין: באיזו מידה מערכת

התגים מתאימה באמת לשני הטקסטים השונים? באיזו מידה עלינו לעצב אותה מחדש – או, לחילופין, להיצמד לקטגוריות הישנות גם (ואולי אפילו דווקא) כאשר הן לא לגמרי מתאימות? לשאלות אלה מימד קונספטואלי יותר מאשר טכני, והתשובות להן, כרגיל, משתנות ממחקר למחקר, ולעתים אף במהלך מחקר אחד. ככל שצוברים נסיון במערכת לומדים כיצד לארגן אותה בהתאם להתלבטויות אלה.

13.24. בדוגמה הנדונה כאן, כך נראה הטקסט השני לאחר שתוייג:

ואמר רבי יהושע בן לוי: בשעה שעלה משה למרום מצאו להקדוש ברוך הוא שהיה קושר כתרים לאותיות. אמר לו: משה, אין שלום בעירך? אמר לפניו: כלום יש עבד שנותן שלום לרבו? אמר לו: היה לך לעזרני. מיד אמר לו: 'ועתה יגדל נא כח ה' כאשר דברת'.

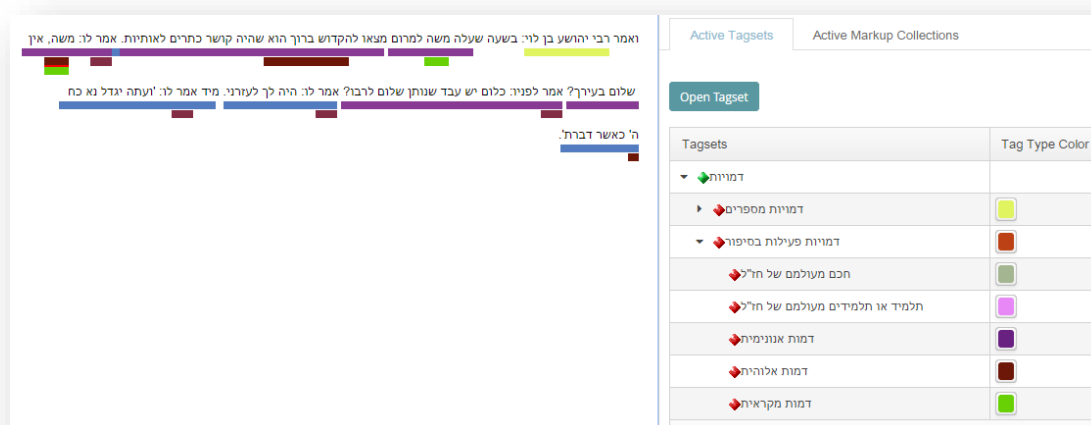
13.25. לא כאן המקום לעסוק בפרטי הניתוח המוצע. יצוין רק כי הסיפור מציב אתגר מיוחד בפני הנסיון לאפיין

את האירועים המתרחשים בו, שכן האירוע המרכזי – לפחות מזווית מבטו של הקדוש ברוך הוא – הוא שתיקתו של משה, שאינה מתועדת בטקסט אלא על דרך השלילה בלבד; במונחים נראטולוגיים, שתיקה זו היא בבחינת 'eventful non-event'. גם אם נוסיף כעת תג מסוג חדש למערכת התגים, עדיין יהיה עלינו להכריע כיצד לקשור אותו אל הטקסט. כדי לפתור את האתגר בחרתי להימנע הפעם מיצירת תג חדש, ובמקום זאת העדפתי

לסמן כ'אירוע בעל משמעות מיוחדת' את הרווח שלפני דברי הקב"ה הראשונים בסיפור (תג כחול מסומן בקו אדום דק בתמונה לעיל).

13.26. תיוג הסיפור השני מדגים היטב סוגיה נוספת, עקרונית ומהותית ל-CATMA. פרשנות סבירה לגמרי

לסיפור תזזה בו יסודות המפקיעים את משה בהדרגה מהיות 'דמות מקראית', והופכים אותו כמעט ל'דמות אלוהית'. לאו דווקא בגלל העליה למרום (המתוארת גם כאן כמבלי משים), אלא בעיקר הודות לביטוי המופגן של הרעיון לפיו משה יכול וצריך לעזור לבוראו; משה מתחיל את הסיפור כבן-אנוש זהיר, ומסיים אותו כשותף לאלוהיו. אם נבקש לתעד קריאה זו בשכבת התיוג, נוכל, דרך משל, להוסיף להופעתו השניה של השם 'משה' את התג 'דמות אלוהית' (חום), בנוסף לתג 'דמות מקראית' (ירוק).



13.27. דוגמא זו מבהירה בקיצור נמרץ למה התכוונו יוצרי CATMA כשתיארו אותה כאמצעי אידיאלי ל-

'**undogmatic reading**' (משחק המילים על אודות הכלב והחתול אוכל, למרבה הצער, בתרגומו לעברית): הואיל והמשתמש רשאי לקבוע את הקטגוריות המנחות אותו כרצונו, הואיל ודבר אינו מונע ממנו מלתעד כל קריאה העולה על דעתו (לרבות קריאות סותרות), החופש הפרשני שלו – העולה לעולם על זה של המחשב – זוכה להכרה מלאה, והופך לשכבת מידע סובייקטיבית, לגיטימית לגמרי, בין שלל השכבות התורמות לניתוח הטקסט.

13.28. כעת, מכל מקום, שני הטקסטים המרכיבים את הקורפוס המודגם כאן מתוייגים, בתום תהליך ידני.

קוראים רבים עשויים לתהות כאן מה ערכו של התיוג, ומה אפשר להפיק ממנו; לשאלה זו מציעה CATMA שלוש תשובות עיקריות.

13.28.1. ראשית, יש ערך בפעולת התיוג כשהיא לעצמה. היא אמנם מגבילה ודווקנית יותר מקריאה חופשית רופפת, משום שהחוקר נדרש לקשר את קריאתו לאלמנטים קונקרטיים בטקסט; אך דרישה זו מעמיקה את מודעותו להכרעות פרשניות מסוימות, ומחדדת את תשומת לבו לצורך בבדיקה ובניסוח מחודש של קטגוריות היסוד המשמשות אותו בעבודתו.

13.28.2. שנית, התבוננות מחודשת בטקסט המתוייג מאפשרת תפיסה טובה יותר שלו; וככל שהטקסט (או הקורפוס) גדול יותר, ככל שמערך התיוג סבוך יותר, כן מתגלה יתרונה של המערכת בניתוח הממצאים. כך, למשל, התיוג הצבעוני מספק עדות יעילה לא רק למציאותן של תופעות מסוימות בטקסט, אלא גם ליחסים ביניהן. בנוסף, הוא מאפשר לעמוד על מקומן במארג הכולל של הטקסט. המערכת שלפנינו מציעה כלי רב-עוצמה המסייע בהתבוננות זו (ומבהיר מדוע עקרון הפרדת השכבות העומד ביסודה משתלם כל כך). אם נפתח במסך ה-Tagger את הלשונית Active Markup Collections הנמצאת באזור הימני של המסך, נוכל 'לשחק' עם השכבות המוצגות על גבי הטקסט (בצידו השמאלי של המסך): הוספה או הסרה של סימן ✓ מן התיבות השונות בעמודת ה-Visible, מאפשרות שליטה מוחלטת בתצוגת התגים, ושינוי מיידי וגמיש שלה בהתאם לכל צורך או רעיון. אם, נניח, מבקש החוקר להתמקד בבחינת היחס בין 'דמויות מספרים' מסוימות ל'אירועים בעלי משמעות מיוחדת', כל שעליו לעשות הוא להגדיר תגים אלה כ-Visible, ולהסתיר באופן זמני את כל שאר התגים לבל ייצרו 'רעש' מיותר על גבי הטקסט (יושם אל לב: הצגת תג מותנה בהגדרת מערכת התגים שאליה הוא שייך כ-Visible. לכן, בדוגמא המובאת כאן, שתי מערכות התגים מסומנות ב-V, אך מתוכן האלמנט הנראה היחיד הוא תג אחד בכל מערכת).

Active Tagsets		Active Markup Collections	
Open Markup Collection		Close	Select all visible
			Deselect all visible
Markup Collections	Tag Type color	Visible	Writable
מרכיבי העלילה בסיפור התלמודי		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
אירועים		<input checked="" type="checkbox"/>	
אירועים מקדמי-עלילה		<input type="checkbox"/>	
אירועי דיבור		<input type="checkbox"/>	
אירועים בעלי משמעות מיוחדת		<input checked="" type="checkbox"/>	
דמויות		<input checked="" type="checkbox"/>	
דמויות מספרים		<input checked="" type="checkbox"/>	
דמויות פעילות בסיפור		<input type="checkbox"/>	

13.28.3. שלישית, משעה ששכבת התיוג מונחת לפנינו, הרי היא עומדת לניתוח אוטומטי ככל שכבת מידע אחרת – ובצד כל שכבת מידע אחרת, לרבות הטקסט לגופו. הפרקים הבאים יוקדשו לסקירת כמה נתיבים שימושיים בניתוח הטקסט המתוייג.

14.1. ניתוח אוטומטי של טקסט או קורפוס מתיוג אינו שונה מבחינה טכנית מניתוח אוטומטי של טקסט או קורפוס שלא תיוג, כפי שהוצג בהרחבה לעיל. אולם מכל בחינה אחרת הוא שונה בתכלית: לאחר שתיעדנו באמצעות התיוג כיצד הטקסט נקרא מזווית מבטו של החוקר, אנו מבקשים מהמערכת להתייחס לקריאה זו כאל שכבת מידע 'לגיטימית' נוספת על גבי הטקסט; שכבה שאפשר לנתח, לבחון, ולעמת עם העולה משכבות אחרות.

14.2. כדי לטעון טקסט במסך ה-Analyzer נלחץ על כפתור ה-

Analyze Document הנמצא מתחת לטקסט הרלוונטי במסך ה-

Tagger. יושם אל לב: כל עוד לא נטען טקסט למסך ה-Analyzer, מסך

A rectangular button with a light blue gradient and a thin white border. The text "Analyze Document" is centered in a white, sans-serif font.

זה ישאר ללא פעילות. במילים אחרות, אם נעבור אליו כעת בלי לטעון טקסט, לא נראה דבר.

14.3. כזכור, אפשרות נוספת לטעון את הטקסט לניתוח נתונה ישירות ממסך ה-Repository Manager באחת משתי

דרכים: (א) אם מדובר בקורפוס, באמצעות לחיצה על כפתור Analyze Corpus הנמצא מתחת לטבלת

הקורפוסים; (ב) אם בטקסט בודד מדובר, באמצעות לחיצה על כפתור More actions הנמצא תחת טבלת

המסמכים, ולחיצה נוספת על Analyze Document בין שלל האפשרויות הנפתחות שם. הנוקטים בדרך זו

צריכים להקפיד ללחוץ תחילה על שם הטקסט, ולאחר מכן לסמן את אסופת התיוג (Markup collection)

הרלוונטית – ורק אז ללחוץ על Analyze Document. שאם לא כן, הטקסט יטען לבדו, בלי שכבת התיוג.

14.4. כלי שימושי במיוחד במסך ה-Analyzer נועד דווקא להפוך את התיוג לאוטומטי-למחצה: אנו יכולים להורות

למערכת לתייג בבת אחת – בכל הטקסט, בכל הקורפוס – סדרת ממצאים שעלו בעקבות שאילתא, ובכך להעמיק

ולהעשיר במהירות את התיוג הידני. לתכונה זו תכליות מגוונות, ולהלן נעמוד רק על שתי העיקריות שבהן.

14.5. הדוגמה המוצעת כאן מספקת מקרה מבחן נוח לכך. כבר ראינו שאחד האירועים השכיחים ביותר בסיפורים

התלמודיים שלפנינו (וגם ברבים אחרים) הוא ההיגד; האמירה. במידה לא מבוטלת, ומסיבות שלא יידונו כאן,

הסיפורת התלמודית נוטה להיות סיפורת של אמירות וחילופי דברים, יותר מאשר סיפורת של פעולות פיזיות. ואולם, לאירועים אלה סמן לשוני מובהק – הופעות שונות של מילים משורש א.מ.ר. בהתחשב בעובדה זו, שמא נוכל לתייג את כלל האירועים הללו בבת אחת?

14.6. כדי לעשות זאת, ניגש ראשית למסך ה- Tag Type Manager, ותחת התג 'אירועים' נסיף את תג-המשנה: 'אירועי-דיבור'. כעת, כדי 'ללכוד' את כלל אירועי-הדיבור בקורפוס, נפנה למסך ה- Repository Manager, נסמן את הקורפוס הרלוונטי, ונלחץ על כפתור Analyze Corpus.

14.7. במסך ה- Analyzer נבנה שאילתא מבוססת דומות (by grade of similarity), ונבקש מהמחשב להפיק רשימה של כל המילים הדומות למילה 'אמר' ברמת דומות של 70%. התוצאות שהתקבלו כוללות 17 הופעות של 'אמר', הופעה אחת של 'אמר', והופעה אחת נוספת של 'ואמר' (סך הכל 19 מופעים).

Result by phrase			Result by markup
Phrase	Frequency	Visible in Kwic	
▶ אמר	17	<input checked="" type="checkbox"/>	
▶ אמרו	1	<input checked="" type="checkbox"/>	
▶ ואמר	1	<input checked="" type="checkbox"/>	

14.8. נסמן ✓ בכל הריבועים שבעמודת Visible in Kwic, ונקבל את כלל הופעותיהן של מילים אלה בהקשרן הטקסטואלי המדויק בטבלה שבחלקו הימני של המסך:

Document/Collection	Left Context	Keyword	Right Context	Start Point	End Point
משה בבית מדרשו של רבי עקיבא	רבי יהודה אמר רב:	אמר		0	3
משה לא אומר שלום לקב"ה	רבי יהושע בן לוי:	ואמר		0	4
משה בבית מדרשו של רבי עקיבא	רב: בשעה שעלה משה	אמר	אמר רבי יהודה	13	16
משה בבית מדרשו של רבי עקיבא	לפניו: רבונו של עולם	אמר	שיושב וקושר כתרים לאותיות.	89	92
משה לא אומר שלום לקב"ה	לוי: משה, אין	אמר	שהיה קושר כתרים לאותיות.	89	92
משה לא אומר שלום לקב"ה	לפניו: כלום יש עבד	אמר	, אין שלום בעירך?	118	121
משה בבית מדרשו של רבי עקיבא	לוי אדם אחד יש שעתידי	אמר	מי מעכב על ידך?	131	134
משה לא אומר שלום לקב"ה	לוי: היה לך לעורני	אמר	עבד שנותן שלום לרבו?	158	161
משה לא אומר שלום לקב"ה	לוי: 'עתה יגדל	אמר	היה לך לעורני. מיד	185	188
משה בבית מדרשו של רבי עקיבא	לפניו: רבונו של עולם	אמר	תילין תילין של הלכות.	245	248
משה בבית מדרשו של רבי עקיבא	לוי: חזור לאחורך.	אמר	עולם, הראהו לי!	281	284
משה בבית מדרשו של רבי עקיבא	לוי תלמידיו: רבי,	אמר	. כיון שהגיע לדבר אחד	384	388
משה בבית מדרשו של רבי עקיבא	לוי: הלכה למשה מסיני	אמר	רבי, מנין לך?	415	418
משה בבית מדרשו של רבי עקיבא	לפניו: רבונו של עולם	אמר	לפני הקדוש ברוך הוא,	484	487
משה בבית מדרשו של רבי עקיבא	לוי: שמעו, רב	אמר	ונתן חורב על ידי?	547	550

14.9. כזכור, הופעותיהן השונות של המילים שחיפשנו בשאילתא נחשבות בטבלה זו ל- Keywords, ומופיעות במרכז

הטבלה. משני עבריהן אנו רואים את ההקשר המילולי שבו כל אחת מהן מופיעה, מסודר מימין (Right

Context) לשמאל (Left Context), בהתאם לשפתו העברית של הטקסט. בעמודה השמאלית ביותר נמצא את

שם הטקסט שבו כל מופע מתועד; ובשתי העמודות הימניות ביותר נמצא נתונים מספריים – אמנם, לא מאוד

מעניינים אולי מצד עצמם – באשר לנקודת ההתחלה ולנקודת הסיום של כל מופע בטקסט (המערכת סופרת כל

תו בנפרד, ולכן, למשל, הופעתה השניה של המילה 'אמר' בסיפור 'משה בבית מדרשו של רבי עקיבא', מתחילה

בנקודה 13 ומסתיימת בנקודה 16).

14.10. נוכל לסדר את הטבלה בהתאם לצרכינו, לפי כל אחד מהפרמטרים המצוינים בשורה העליונה: בחלוקה

למסמכים, לפי הסדר האלפביתי של ההקשרים ושל מילת המפתח עצמה, ואף לפי מקום המילים ברצף הטקסט;

למעבר בין בחירות אלה נלחץ לחיצה אחת או שתיים על הפרמטר המבוקש בראש הטבלה.

14.11. בדוגמה הנדונה כאן, סודרה הטבלה בהתאם למיקום מילות המפתח ברצף הטקסט מתחילתו עד סופו

(כלומר, העמודה הקובעת את הסדר היא עמודת ה- Start point, ולכן מופיע בצידה משולש קטן). במבט חטוף

נקל לראות כי שלושת המופעים הראשונים של המילה 'אמר' אינם בבחינת 'אירועי-דיבור', שכן תפקידם להציג

את דברי המספרים, ולא את דברי הדמויות הפעילות בסיפורים. לעומתם, כל יתר המופעים יכולים להיחשב,

ובדין, לאירועי-דיבור סיפוריים.

14.12. לאור

זאת נסמן את

כלל שאר

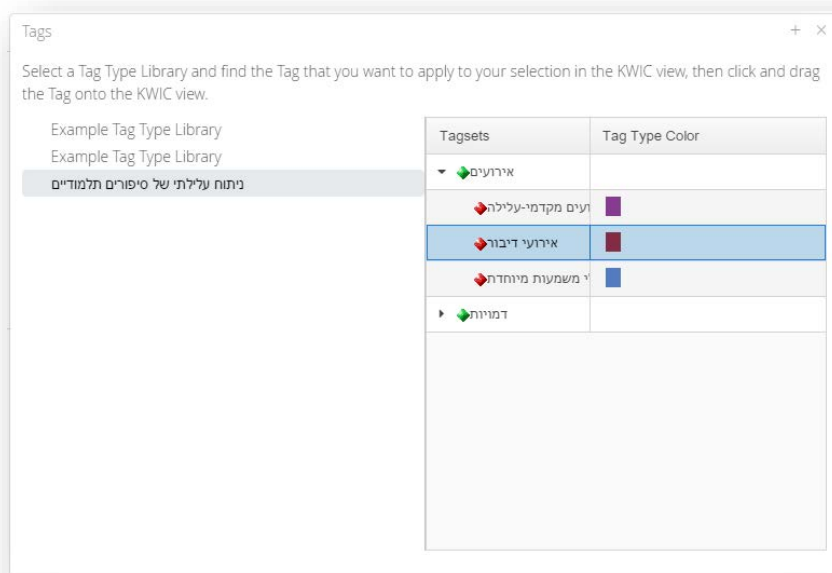
המופעים, כך:

Document/Collection	Left Context	Keyword	Right Context	Start Point	End Point
משה לא אומר שלום לקב"ה	רבי יהושע בן לוי:	ואמר		0	4
משה בבית מדרשו של רבי עקיבא	רבי יהודה אמר רב:	אמר		0	3
משה בבית מדרשו של רבי עקיבא	רבי: בשעה שעלה משה	אמר	אמר רב יהודה	13	16
משה בבית מדרשו של רבי עקיבא	לפניו: רבונו של עולם	אמר	שישב וקושר כתרים לאותיות.	89	92
משה לא אומר שלום לקב"ה	לה משה, אין	אמר	שהיה קושר כתרים לאותיות.	89	92
משה לא אומר שלום לקב"ה	לפניו: כלום יש עבד	אמר	, אין שלום בעירך?	118	121
משה בבית מדרשו של רבי עקיבא	לו אדם אחד יש שעתידי	אמר	מי מעבב על ידך?	131	134
משה לא אומר שלום לקב"ה	לה היה כך לעזוני	אמר	עבד שנותן שלום לרבו?	158	161
משה לא אומר שלום לקב"ה	לה: 'ועתה יגדל	אמר	היה כך לעזוני. מיד	185	188
משה בבית מדרשו של רבי עקיבא	לפניו: רבונו של עולם	אמר	חילין חילין של הלכות.	245	248
משה בבית מדרשו של רבי עקיבא	לה: חזור לאחורך.	אמר	עולם, הראהו לי!	281	284
משה בבית מדרשו של רבי עקיבא	לו תלמידיו: רבי,	אמר	. כיון שהגיע לדבר אחד	384	388
משה בבית מדרשו של רבי עקיבא	לה: הלכה למשה מסיני	אמר	רבי, מנין קי?	415	418
משה בבית מדרשו של רבי עקיבא	לפניו: רבונו של עולם	אמר	לפני הקדוש ברוך הוא,	484	487
משה בבית מדרשו של רבי עקיבא	לה: שחוק, כך	אמר	נותן תורה על ידי?	547	550
משה בבית מדרשו של רבי עקיבא	לפניו: רבונו של עולם	אמר	כך עלה במחשבה לפני.	581	584

14.13. כעת נלחץ על כפתור Tag selected results הנמצא בפינה הימנית-תחתונה של המסך.

14.14. בחלון שנפתח, נבחר את ספריית התגים ואת מערכת התגים הרלוונטיות, ולבסוף את התג שאותו נבקש

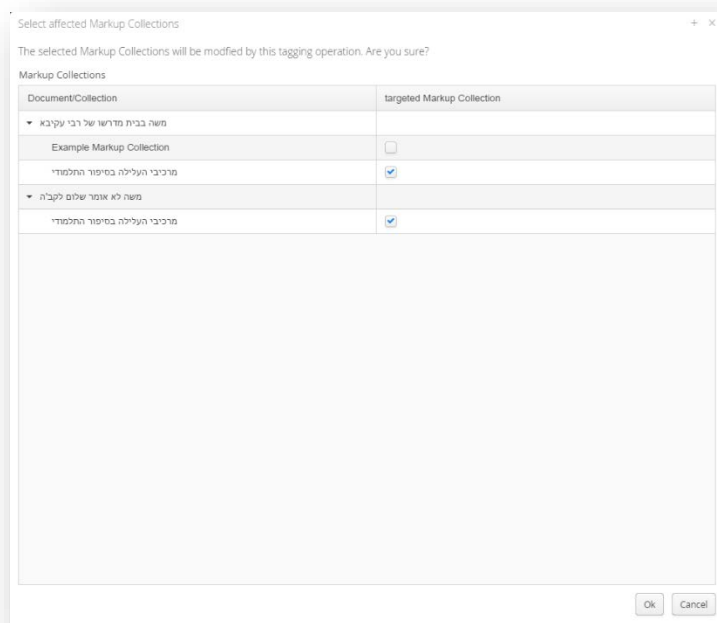
לצרף לכל המילים שסומנו בטבלת ה-Kwic:



14.15. עתה 'נגרור' עם העכבר את התג המבוקש אל עבר תיבת ה-Kwic. המערכת תבקש מאיתנו לאשר את

אסופת התיוג שבה יישמר התיוג החדש – במקרה זה, שתי אסופות תיוג (בעלות שם זהה, בהתאם לקביעתנו

לעיל) המשויכות לשני טקסטים שונים בקורפוס.



14.16. לאחר לחיצה על

כפתור Ok, נקבל בפינה

הימנית-תחתונה של

המסך הודעה לפיה 16

תיוגים חדשים נוספו

לאסופות תיוג אלה.

14.17. כשנפתח בפעם הבאה את הטקסטים במסך ה-Tagger, ונבקש מהמערכת (בלשונית ה-Active

Markup Collections) להציג את כל התיוגים, נראה את תוצאת התיוג האוטומטי-למחצה שתואר כאן. אפשר לדמיין בנקל את מידת החסכון בזמן הצפויה בעקבות שימוש באפשרות זו, ככל שמדובר בטקסטים גדולים או בקורפוסים אדירי מימדים.

14.18. בדרך דומה אפשר להשיג תכלית נוספת, שימושית מאוד במחקרים חישוביים רבים – יצירת אבחנה בין

תופעות שונות שפני השטח הטקסטואליים שלהן זהים. כך:

14.19. הבה נניח שאנו חוקרים רומן רחב-יריעה ומעוניינים לעמוד על מקום הדמויות השונות בתוכו. לא פעם

נגלה כי חיפוש פשוט של תפוצת שמות הדמויות מניב תוצאות חלקיות ואף לא-מדויקות, למשל כאשר חלק מהשמות משמשים גם כמילים עצמאיות בעלות מובן שונה (דוגמאות בעברית: שמות כמו 'אשר', 'אילן', 'יעל', 'ירדן', 'ישראל', ועוד). תיוג אוטומטי-למחצה מספק במקרים אלה פתרון מצוין: העברת המילה כפולת-המשמעות לחלון ה-Kwic תאפשר לסקור בקלות את כל מופעיה, ולסמן במהירות רבה את המופעים הרלוונטיים מתוכה – ורק אותם – בתג ייעודי (בדוגמה זו, למשל, בתג 'שם'). כאשר נבקש לאחר מכן לבחון את תפוצת השם הרלוונטי, נוכל להגביל את תוצאות החיפוש המילולי לאלה המסומנות בתג המבחין שיצרנו (באמצעות יצירת שאילתא מורכבת; וראו להלן) – וכך לקבל תמונה טובה ומדויקת יותר.

14.20. שיטה זו יכולה להתאים כמובן לתיוג כל ממצא שאותר באמצעות כל שאילתא, תהא אשר תהא. לפיכך

אפשר להחיל אותה גם על שאילתות ברמת מורכבות משתנה, לרבות כאלה המתמקדות בשכבת התיוג יותר מברשימת המילים. שאילתות כאלה תוצגנה בפרקים הבאים.

14.21. את חטיבת פרקי התיוג נבקש לחתום בהערה עקרונית אחרונה. תיוג פרשני שיטתי, מסועף וממושך של

טקסט או של קורפוס גדול עשוי להעמיד בפני החוקר שכבת metadata עצומה בהיקפה, עשירה באבחנותיה ומשוכללת עד מאוד. כדי להפיק ממנה את המירב נדרש – שוב – מאמץ פרשני ניכר; נדרש גם אורך-רוח. הממצאים לעולם אינם 'מדברים' באופן עצמאי, בלי שנדוּבב אותם בדרך מן הדרכים. פעולות הניתוח המתוארות בהמשך, אם כן, אינן יותר מאשר הצעות לטיפול ראשוני בחומר המתוייג; ואידך זיל גמור.

15. ניתוח אוטומטי של טקסט או של קורפוס מתויג (א): שאלות פשוטות

15.1. עד כה התמקדנו בניתוח אוטומטי של טקסט על בסיס רשימת המילים ממנה הוא עשוי, או חיתוכים שונים שלה – כלומר, על בסיס שכבת ה- data. בתחילת הדרך היה זה ניתוח פשוט ביותר, שהופק באמצעות לחיצה על כפתור Wordlist. בשלב מתקדם יותר היו אלה שאלות פשוטות ומורכבות שהופקו בעזרת בונה-השאלות. בפרק הנוכחי נלמד כיצד לרתום את בונה-השאלות לטיפול בשכבת ה- metadata – אסופת התיג (Markup collection) שיצרנו קודם – וכן לניתוח היחסים בין השכבות השונות.

15.2. בשונה משכבת ה- data, הניתנת לסקירה מהירה בעזרת כפתור ה- Wordlist, אין במערכת (לפי שעה) אפשרות לסקירה מהירה של שכבת ה- metadata באמצעות הצגה אוטומטית של כל התגים במסך ה- Analyzer (מי שמבקש התרשמות מהירה כזו יכול להשיגה ישירות במסך ה- Tagger, כפי שהוסבר לעיל). לפיכך, נקודת המוצא לכל ניתוח של שכבה זו היא בניית שאלתא מתאימה. אנו נמשיך גם כאן לבנות את השאלות באמצעות בונה-השאלות, ולא בעזרת ניסוח ישיר של השאלתא בשורת הפקודות שבחלקו העליון של המסך (המתעניינים באפשרות האלטרנטיבית ימצאו את מבוקשם בנספח).

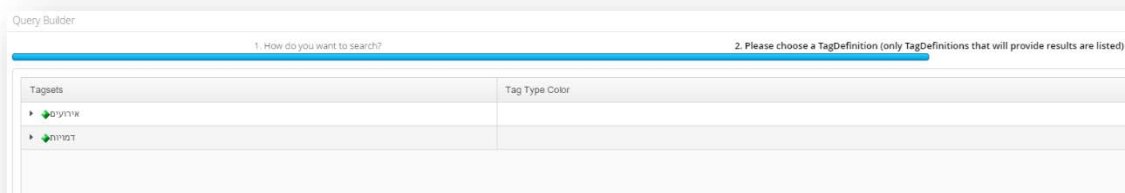
15.3. נלחץ, אם כן, על כפתור ה- Query Builder, ונקבל את החלון הראשון של בונה-השאלות, שבו אנו מתבקשים להגדיר את צורת החיפוש המבוקשת ('How do you want to search?'). נסמן 'by Tag' (שאר האפשרויות מיועדות לחיפוש בטקסט עצמו, כלומר, בשכבת ה- data). נלחץ על כפתור Next הנמצא בפינה

הימנית-

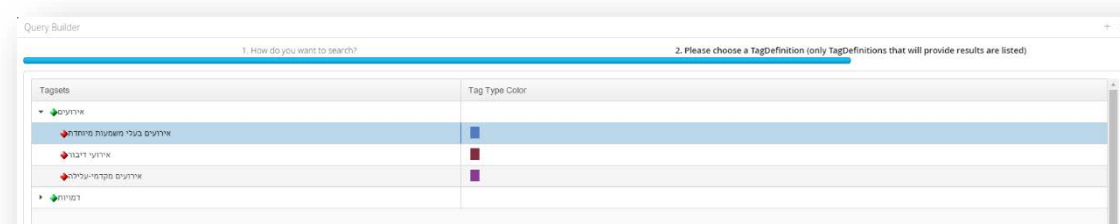
תחתונה של

המסך.

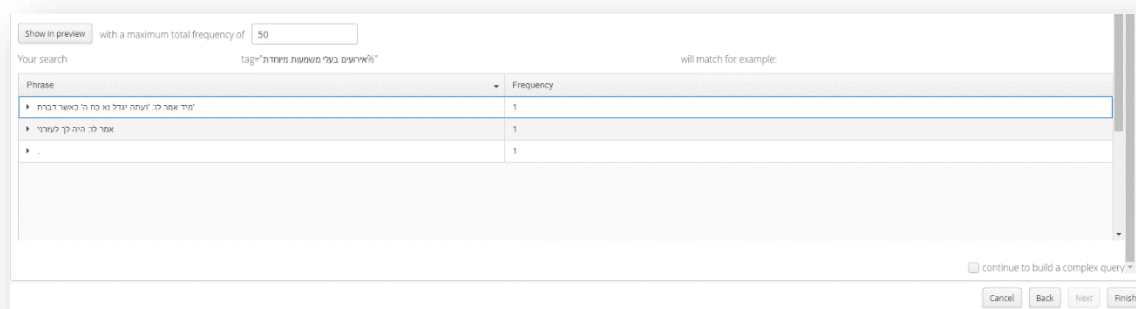
15.4. בחלון הבא המערכת מציגה את מערכות התגים (Tagsets) שבהן השתמשנו עד כה בתיוג הטקסט, ומבקשת מאיתנו לבחור את התג שנרצה להעמיד במוקד השאילתא (שאלה זו מנוסחת מעל הקו הכחול הדק שבראש המסך: Please choose a Tagdefinition).



15.5. יושם אל לב: מערכות התגים, כפי שכבר הובהר קודם, אינן 'תג' בפני עצמו; הן אינן חלק משכבת ה-metadata, אלא רק אמצעי וירטואלי לארגון התגים בספריית התגים המופשטת, תחת מכנה משותף שנקבע על ידי המשתמש. לפיכך, אם נסמן אותן כעת, לא נקבל שום תוצאה. לעומת זאת, לחיצה על המשולש הקטן שלצד שמות מערכות התגים תפרוס לפנינו את התגים שבהם השתמשנו עד כה.



15.6. סימון אחד מהם (למשל, 'אירועים בעלי משמעות מיוחדת'), ולחיצה על כפתור Show in preview (כזכור, תנאי הכרחי לכל שלב בכל שאילתא המנוסחת באמצעות בונה-השאילתות), תציג בפנינו את כל קטעי הטקסט שתויגו בתג המבוקש (במקרה זה, שלושה מקטעים בסיפור 'משה לא אומר שלום לקב"ה'; שניים מהם פעולות דיבור, ואחד סומן כשתיקה).



15.7. לחיצה על כפתור Finish תחזיר אותנו למסך ה-Analyzer, והפעם תוך הצגת נוסח השאילתא (בתיבת ה-

Query) וממצאי השאילתא.

The screenshot shows the Kwic interface. At the top, there is a 'Query' section with a text input field containing the query: `tag="אירועים בעלי משמעות מיוחדת"`. Below the input field are two buttons: 'Query Builder' and 'Wordlist', followed by a help icon. Below this is a tabbed interface with two tabs: 'Result by phrase' (selected) and 'Result by markup'. The 'Result by phrase' tab displays a table with three columns: 'Phrase', 'Frequency', and 'Visible in Kwic'. The table contains three rows of results:

Phrase	Frequency	Visible in Kwic
מִיד אָמַר לוֹ: 'וְעַתָּה יִגְדֹּל נָא כַח ה'	1	<input type="checkbox"/>
.	1	<input type="checkbox"/>
אָמַר לוֹ: הִיָּה לָךְ לְעוֹרָנִי	1	<input type="checkbox"/>

15.8. יושם אל לב, כי ממצאי השאילתא יכולים להיות מסודרים בשתי תצורות שונות – שתי זוויות מבט, אפשר לומר

– המאפשרות מעבר מהיר מאחת לשנייה באמצעות הלשוניות שמעל טבלת הממצאים:

15.8.1. **Result by phrase** – בתצורה זו, 'ממצא' הוא כל מקטע טקסט שסומן בתג שנבחר בבניית

השאילתא. מקטעים אלה הם העומדים במוקד, כלומר, מבטנו מופנה כעת אל הטקסט (שכבת ה- data)

כפי שהוא משתקף דרך הפרויזמה של התיוג, ולא אל התג לגופו (בדוגמא זו: שלושה מקטעי טקסט שסומנו

בתג 'אירועים בעלי משמעות מיוחדת').

15.8.1.1. אם נלחץ על המשולש הקטן שמשמאל לכל ממצא, נקבל את שם הטקסט שבו אות. בהמשך

נלמד כיצד להפיק תועלת ממידע זה.

Result by phrase			Result by markup
Phrase	Frequency	Visible in Kwic	
▼ מִיד אָמַר לוֹ: 'וְעַתָּה יִגְדֹּל נָא כַח ה' כְּאִשֶּׁר דִּבֶּרְתָּ	1	<input type="checkbox"/>	
מִשֶּׁה לֹא אָמַר שְׁלוֹם לְקַבִּי	1		
▼ .	1	<input type="checkbox"/>	
מִשֶּׁה לֹא אָמַר שְׁלוֹם לְקַבִּי	1		
▼ אָמַר לוֹ: הִיָּה לָךְ לְעוֹרָנִי	1	<input type="checkbox"/>	
מִשֶּׁה לֹא אָמַר שְׁלוֹם לְקַבִּי	1		

15.8.2. **Result by markup** – בתצורה זו, 'ממצא' הוא כל תג שנבחר בתהליך בניית השאילתא. תגים אלה

הם העומדים במוקד, כלומר, מבטנו מופנה כעת אל אסופת התיוג (שכבת ה-metadata), ורק בעקיפין

אל הטקסט עצמו (בדוגמא זו: שלוש היקרויות של התג 'אירועים בעלי משמעות מיוחדת').

Result by phrase Result by markup		
Tag Type Definition	Frequency	Visible in Kwic
▶ אירועים בעלי משמעות מיוחדת/	3	<input type="checkbox"/>

15.8.2.1. אם נלחץ על המשולש הקטן שמשמאל לתג המוצג ברשימת הממצאים, נקבל תחילה את שם

הטקסט/ים שבו/בהם התג סומן; אם נלחץ גם על המשולש שלידו, נראה את שם אסופת התיוג שבה

נשמר התיוג הרלוונטי; לחיצה נוספת על המשולש שלידו מציינת בזו אחר זו את ההיקרויות השונות

של התיוג – אך לא את הטקסט המתויג. בהמשך נראה כיצד אפשר להפיק תועלת ממידע זה; בשלב

הנוכחי די בכך כדי להדגים שבתצורה זו לא הטקסט עומד במרכז, אלא, נאמר, האפיון הפרשני שלו.

Result by phrase Result by markup		
Tag Type Definition	Frequency	Visible in Kwic
▼ אירועים בעלי משמעות מיוחדת/	3	<input checked="" type="checkbox"/>
▼ משה לא אומר שלום לקב"ה	3	<input type="checkbox"/>
▼ מרכיבי העלילה בסיפור התלמודי	3	<input type="checkbox"/>
אירועים בעלי משמעות מיוחדת/	1	<input type="checkbox"/>
אירועים בעלי משמעות מיוחדת/	1	<input type="checkbox"/>
אירועים בעלי משמעות מיוחדת/	1	<input type="checkbox"/>

15.9. ההבדל בין שתי התצורות בא לידי ביטוי גם באפשרויות הויזואליזציה שלהן במסך ה-Visualizer:

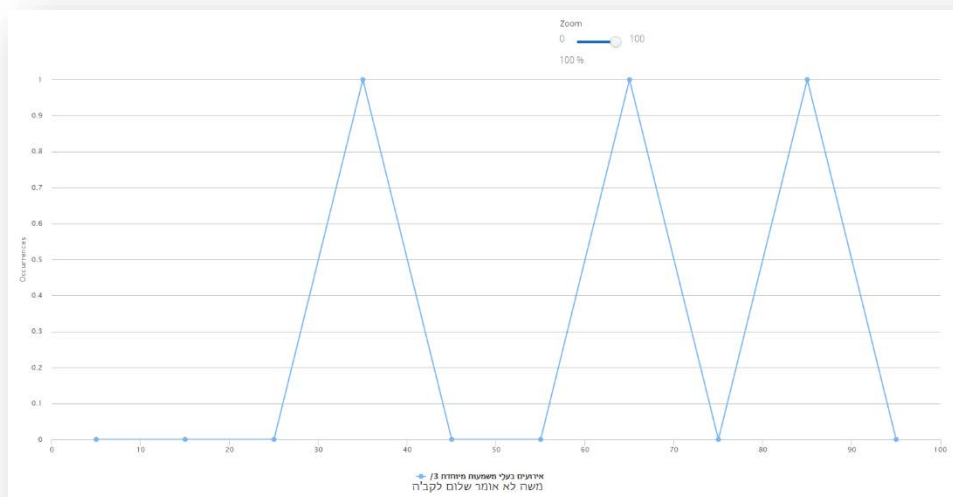
15.9.1. ראשית, גרף מסוג 'עץ כפול' (Double Tree) אפשרי רק בתצורת Result by phrase, משום

שביסודו הוא מבוסס על ניתוח הטקסט עצמו; הואיל ולתגים אין נוכחות של ממש בטקסט כשהוא לבדו,

הם אינם נתונים בהקשר מן הסוג שגרף זה מאתר וממפה.

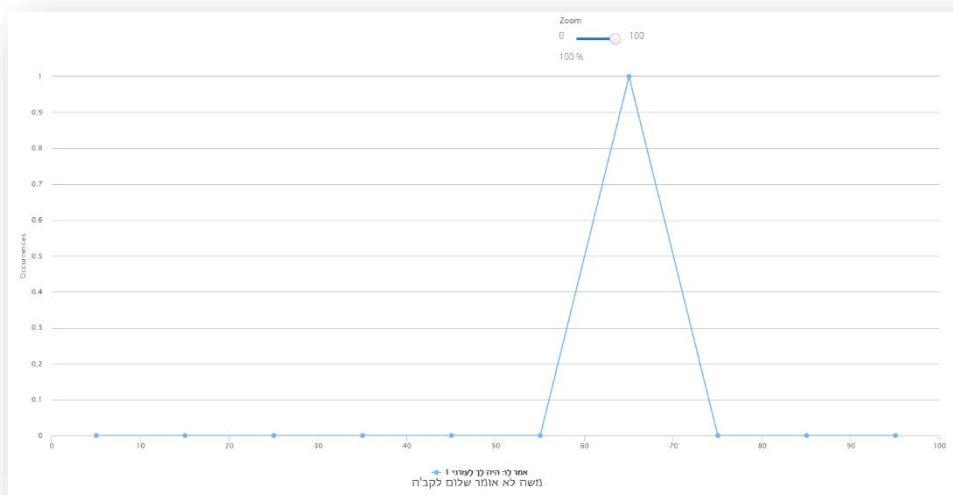
15.9.2. שנית, אם נבקש מהמערכת למקם את הממצאים בגרף תפוצה (Distribution Analysis), נמצא כי

סימון תג אחד בתצורת Result by markup יפיק גרף הכולל את כל מופעי התג בשכבת ה-metadata:

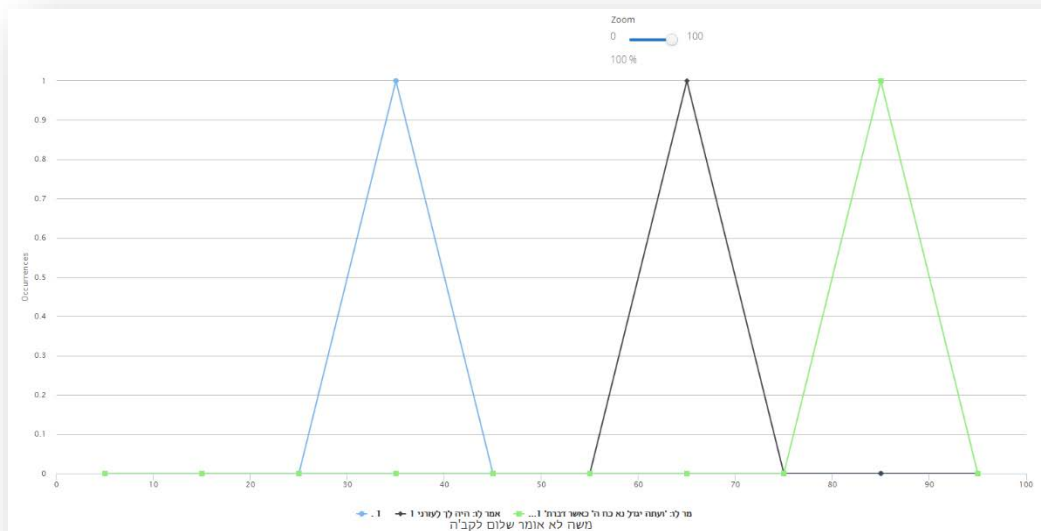


ולעומת זאת, סימון ממצא אחד בתצורת Result by phrase יפיק גרף הכולל אך ורק את מופעיו של

מקטע טקסטואלי זה בשכבת ה-data (במקרה שלפנינו, מופע בודד):



גם בתצורה זו נוכל אמנם להפיק גרף דומה במראהו לגרף הראשון, אך לשם כך יהיה עלינו לסמן בזה אחר זה את כל מקטעי הטקסט שאותרו בזיקתם לתג הרלוונטי בשאילתא (מומלץ ללחוץ על כפתור הפקת הגרף לאחר סימון של כל ממצא בפני עצמו על מנת שהמערכת תייצג כל מקטע בקו עצמאי ובצבע שונה על פני הגרף. לעומת זאת, אם נסמן את כל הממצאים ואז נלחץ על כפתור הפקת הגרף, נקבל קו אחד המייצג את כולם). כך נקבל גרף הכולל את כל היקרויות התג המבוקש – אך דרך הפריזמה של הממצא הטקסטואלי, ולא של הקטגוריה המופשטת המגולמת בתג עצמו:



15.10. הבדלים אלה עשויים להיראות זניחים כעת; אולם מודעות להם עתידה לאפשר מהלכי ניתוח עשירים

יותר בשלבים מתקדמים של עבודה עם המערכת. ככל שנכלול במחקרנו טקסטים רבים יותר, ומערכות תגים מסועפות יותר, כך פונקציות אלה, לגווניהן השונים, תהפוכנה שימושיות יותר.

15.11. בצד כל האמור לעיל, שתי התצורות שוות זו לזו בפונקציה מרכזית אחת: אם נסמן \sqrt בתיבות שבעמודת

ה- Visible in Kwic נקבל, הלכה למעשה, את אותן תוצאות בדיוק בטבלת ה- Kwic (Keywords in

context) שבצידו הימני של המסך. זאת, משום שטבלה זו – בהגדרתה – מתמקדת בהכרח בטקסט עצמו (כלומר, בשכבת ה-data).

Result by phrase			Result by keyword					
Phrase	Frequency	Visible in Kwic	Document/Collection	Left Context	Keyword	Right Context	Start Point	End Point
למד אמר לו: לעולם ידור נא נה רז כאשר דגורל	1	<input checked="" type="checkbox"/>	דגורל לא אמר שאלו חלקים	דגורל לא אמר שאלו חלקים	למד אמר לו: לעולם ידור נא נה רז כאשר דגורל	דגורל לא אמר שאלו חלקים	181	223
•	1	<input checked="" type="checkbox"/>	דגורל לא אמר שאלו חלקים	דגורל לא אמר שאלו חלקים	למד אמר לו: לעולם ידור נא נה רז כאשר דגורל	דגורל לא אמר שאלו חלקים	87	99
אמר לו: הנה כל לעוני	1	<input checked="" type="checkbox"/>	דגורל לא אמר שאלו חלקים	דגורל לא אמר שאלו חלקים	אמר לו: הנה כל לעוני	דגורל לא אמר שאלו חלקים	158	179



Result by phrase			Result by keyword					
Tag Type Definition	Frequency	Visible in kwic	Document/Collection	Left Context	Keyword	Right Context	Tag Type Definition	
• אירועים בעלי משמעות פורמלית	3	<input checked="" type="checkbox"/>	אירועים בעלי משמעות פורמלית	אירועים בעלי משמעות פורמלית	אמר כל היה לו לעוני 1	אמר כל היה לו לעוני 1	אירועים בעלי משמעות פורמלית	
אירועים בעלי משמעות פורמלית			אירועים בעלי משמעות פורמלית	אירועים בעלי משמעות פורמלית	אמר כל היה לו לעוני 1	אמר כל היה לו לעוני 1	אירועים בעלי משמעות פורמלית	
אירועים בעלי משמעות פורמלית			אירועים בעלי משמעות פורמלית	אירועים בעלי משמעות פורמלית	אמר כל היה לו לעוני 1	אמר כל היה לו לעוני 1	אירועים בעלי משמעות פורמלית	

15.12. לקראת סיכום הצגת השאלות הפשוטות בטקסטים מתוייגים, נעמוד על סיפן של שאלות מורכבות

מעט יותר: נראה כיצד אפשר לשלב בגרף אחד את שתי שכבות המידע, כלומר, את ה- data ואת ה- metadata,

במופעהן הפשוטים, הגולמיים כמעט. לשם כך, נחזור אל גרף התפוצה שהופק קודם מתוך תצורת Result by

markup. כדי להוסיף לו אלמנט טקסטואלי פשוט, נשנה את תצורת התוצאות ל- Result by phrase ונלחץ

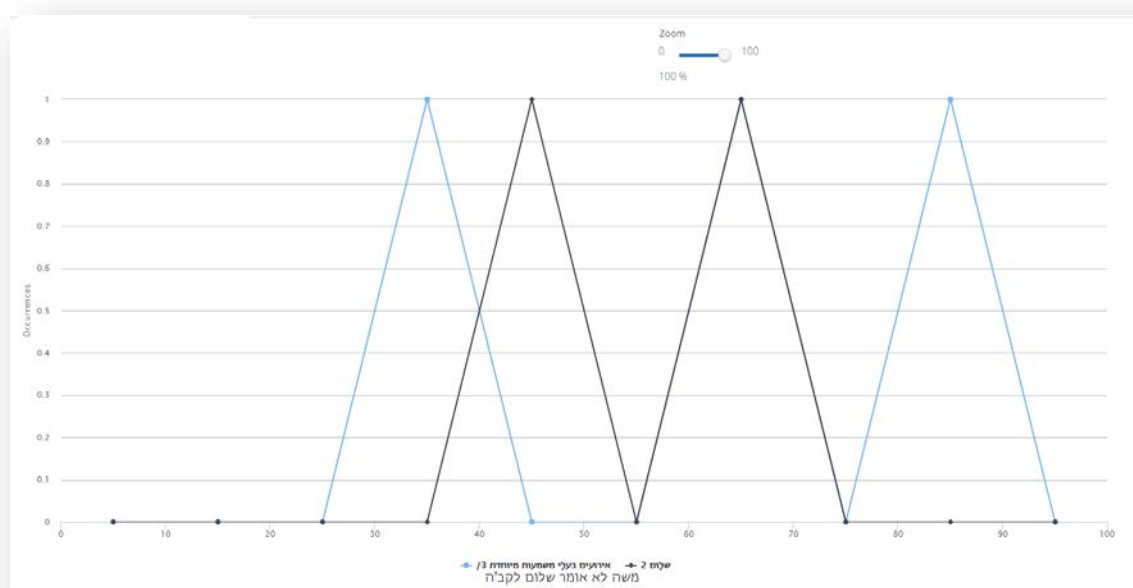
על כפתור ה- Wordlist שבאזור השמאלי-עליון של מסך ה- Analyzer. כעת נקבל שוב את רשימת המילים

המרכיבות את הטקסט, ללא כל זיקה לשאלתא הקודמת (לחיצה על כפתור wordlist מייצרת תמיד שאלתא

פשוטה חדשה ולפיכך משנה את רשימת הממצאים).

15.13. כך למשל בדוגמא הנדונה כאן, סימון המילה 'שלום', ולחיצה על כפתור הפקת הגרף, תצורף את היקרויות

המילה הנבחרת לגרף הקיים, בצד היקרויות התג 'אירועים בעלי משמעות מיוחדת'.



15.14. הנה כי כן, אלמנטים משתי שכבות המידע – אחת אובייקטיבית, אחת סובייקטיבית – ניצבים זה לצד

זה בגרף אחד, ומניחים תשתית אופטימלית לניתוח היחסים ביניהן.

16.1. נניח לעת עתה לויזואליזציה המורכבת שתוארה בסוף הפרק הקודם, ונשוב אל הממצאים בצורתם הטבלאית-טקסטואלית. שאלות פשוטות עשויות להועיל מאוד גם בהשגת תכלית נוספת, שימושית למדי במחקרים רבים: בניית לקסיקון נושאי. בשונה מניתוח אוצר המילים של הטקסט או הקורפוס, המתקיים כולו בשכבת ה- data, אסופת התיוג שיצרנו מאפשרת מיפוי של תופעות שהמכנה המשותף היחיד להן הוא פרשנותנו עצמה, כפי שתועדה בשכבת ה- metadata. מערכת CATMA מסייעת לנו להכין רשימות של מילים, ביטויים ומקטעי-טקסט בהתאם לתיוגם.

16.2. זאת ועוד: לא אחת נמצא את עצמנו מתייגים טקסט מעבר לנדרש על פי מגמתנו הראשונית; פעמים רבות נעדיף לסמן תוך כדי קריאה תופעות שונות שאינן עומדות כעת בהכרח במוקד המחקר שלנו, רק כדי לא 'לאבד' אותן בעתיד. בעוד מחקר 'ידני' עלול לשכוח תופעות אלה, השיטה הפשוטה שתתואר להלן מסייעת למצות את הפוטנציאל הפרשני-מחקרי הטמון במהלכי תיוג עודף.

16.3. בהמשך לדוגמא הנדונה כאן, הבה נניח שברצוננו למפות את כל דמויות המספרים בספרות התלמודית; הבה נניח שברצוננו לדעת כמה סיפורים כל אחד מהם סיפר, מה היו סיפורים אלו, האם יש להם מאפיינים משותפים או אוצר מילים משותף, ועוד. CATMA יכולה לעזור בהוצאת הרעיון מן הכח אל הפועל. לשם כך נפתח תחילה את הקורפוס כולו – שהרי מדובר בשאלה הנוגעת בקורפוס כולו, ולא רק בטקסט הבודד – באמצעות לחיצה על כפתור Analyze Corpus במסך ה- Repository Manager.

16.4. נבנה שאלתא (בתהליך דומה לזה שתואר לעיל) המבקשת מהמערכת להפיק רשימה של כל מה שסומן בתג 'דמויות מספרים'. לאחר לחיצה על Finish נקבל את הרשימה הבאה, תחת הלשונית Result by phrase:

Result by phrase			Result by markup
Phrase	Frequency	Visible in Kwic	
‣ רב יהודה אמר רב	1	<input type="checkbox"/>	
‣ רבי יהושע בן לוי	1	<input type="checkbox"/>	

16.5. רשימה קצרה זו כוללת שני מספרים ; לכל אחד מהם הופעה אחת, משום שכל אחד מהם תועד כמספרו של סיפור אחד משני הסיפורים בקורפוס שלנו – כפי שניתן לראות היטב כאשר לוחצים על המשולש הקטן שליד כל ממצא:

Result by phrase			Result by markup
Phrase	Frequency	Visible in Kwic	
▼ רב יהודה אמר רב	1	<input type="checkbox"/>	
שה בבית מדרשו של רבי עקיבא	1	<input type="checkbox"/>	
▼ רבי יהושע בן לוי	1	<input type="checkbox"/>	
משה לא אומר שלום לקב"ה	1	<input type="checkbox"/>	

16.6. נקל לדמיין, גם כאן, מה תהיה תועלתה של רשימה המתייחסת לקורפוס גדול יותר: יהיה די במבט חטוף כדי לחשוף מספרים דומיננטיים, תמות אופייניות, וכל כיוצא בזה.

16.7. לו עבדנו עם קורפוס גדול יותר, יכולנו בלי כל קושי לפתח את הבדיקה לכיוונים נוספים: יכולנו, למשל, לבנות שאילתא המקבצת את כלל הסיפורים המיוחסים למספר מסוים, ואז לאתר שגרות לשון אופייניות להם (בשכבת ה- data) או תופעות ספרותיות דומות (בהתאם לתיוגן בשכבת ה- metadata). משימות אלה כבר נוגעות בתחומן של שאילתות מורכבות יותר, כפי שנראה בהמשך.



16.8. העברת הרשימות השונות לגליון נתונים, באמצעות לחיצה על הכפתור האחראי לכך מתחת לרשימת הממצאים (Export all query result data as an Excel spreadsheet), תניח יסודות לעיבוד מתקדם נוסף מחוץ ל- CATMA.

16.9. שיטה פשוטה זו לאיתור כל גילומיה הקונקרטיים של קטגוריה מסוימת בטקסט, ותהא אשר תהא אותה קטגוריה, פותחת פתח ליצירת 'לקסיקון' נושאי או מושגי בהתאם לצרכי המחקר, ולהפקת תועלת רבה יותר מעמלנו הפרשני; וכעושר התיוג, כן עושר הממצאים.

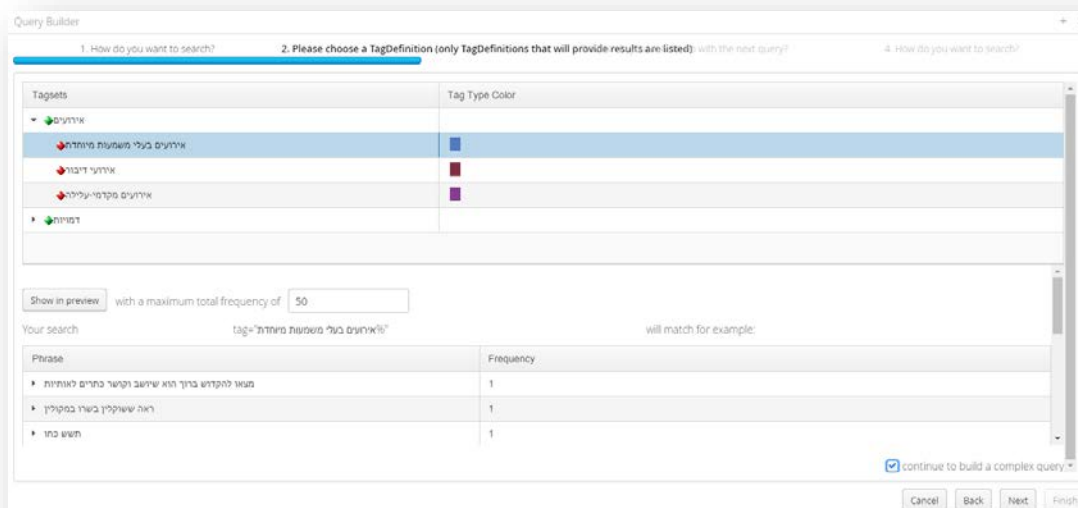
17.1. הבה נניח שברצוננו לבחון את הטקסט או הקורפוס בראי התובנות שהנחו אותנו במהלך התיוג – אך הפעם, בתצורה מורכבת יותר, הכוללת שילוב כמה תגים 'סובייקטיביים' שונים ומידע טקסטואלי 'אובייקטיבי'. לשם כך נעזור באפשרות לבניית שאילתות מורכבות – כלומר, מרובות-שלבים – בבונה-השאילתות (Query Builder) שבמסך ה-Analyzer. בעקרון, ובדומה למה שכבר נאמר לעיל בבניית השאילתות הפשוטות, אין כל הבדל מעשי בין שאילתא מורכבת המתמקדת ב-data, לכזו המתמקדת ב-metadata, או בשילוב של שתיהן יחד; CATMA מעניקה מעמד שוויוני, כביכול, לכל שכבות המידע ולכל אלמנט מדיד בתוכן, יהא אשר יהא מקורו.

17.2. לצורך ההדגמה, ננסה לאתר כעת את כל אותם אירועים שהגדרנו כ'אירועים בעלי משמעות מיוחדת', אך רק בתנאי שיש בהם גם פעולות דיבור; לאחר מכן נבחן את מקומם בקורפוס בהשוואה לכמה ממצאים לשוניים, ונעמוד על היחסים ביניהם.

17.3. הואיל וברצוננו לערוך את הניתוח על הקורפוס כולו, הצעד הראשון הוא לטעון את הקורפוס לניתוח ממסך ה-Repository Manager. ניגש אפוא למסך זה, נסמן את הקורפוס הרצוי ברשימת הקורפוסים שבצדו השמאלי של המסך, נלחץ על כפתור More actions... שתחתיו, ואז נבחר באפשרות Analyze Corpus.

17.4. כעת ניגש לבניית השאילתא במסך ה-Analyzer. לחיצה על כפתור Query Builder תוביל אותנו לחלון הראשון בתהליך בניית השאילתא. נבקש מהמערכת לחפש תוצאות by Tag, ונלחץ על כפתור Next. בחלון הבא נלחץ על המשולשים הקטנים שליד מערכות התגים, ונקבל פירוט של כל התגים שסומנו על גבי הטקסט.

17.5. נסמן את התג הראשון הנחוץ לנו – 'אירועים בעלי משמעות מיוחדת' – ונלחץ על כפתור Show in preview. תוצאות ראשוניות תופענה בטבלה שבחלקו התחתון של החלון.



17.6. כבר במבט שטחי אפשר לראות כי לא כל האירועים שאותרו כוללים בתוכם פעולות דיבור; שני הראשונים,

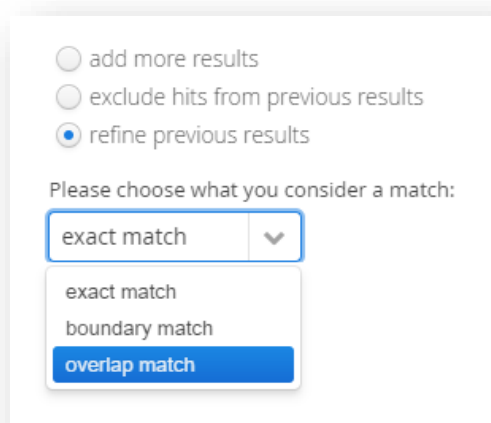
למשל, באים לידי ביטוי במראה ייחודי ומפתיע ("מצאאו...", "ראה..."), ואילו השלישי קשור יותר למצבו הרגשי

או הפיזי של הגיבור ("תשש כוחו"). הממצאים הבאים ניתנים לצפייה בעזרת פס הגלילה הקטן שמימין לטבלת

התוצאות הראשוניות, וחלקם אכן כוללים פעולות דיבור. כדי להתמקד בהם, ובהם בלבד, נסמן $\sqrt{}$ בתיבה שליד

הפקודה continue to build a complex query בפינה הימנית-תחתונה של החלון, כך שלחיצה על כפתור

Next תוביל אותנו לפתיחת שלב חיפוש נוסף, שבעזרתו נשפר את התוצאות.



17.7. בחלון המעבר בין שלבי השאילתא נבקש מהמערכת

לאפיין מחדש את התוצאות הקודמות (refine

previous results), ונבחר באפשרות המתנה אפיון

זה בהתאמה ברמה של חפיפה (overlap match)

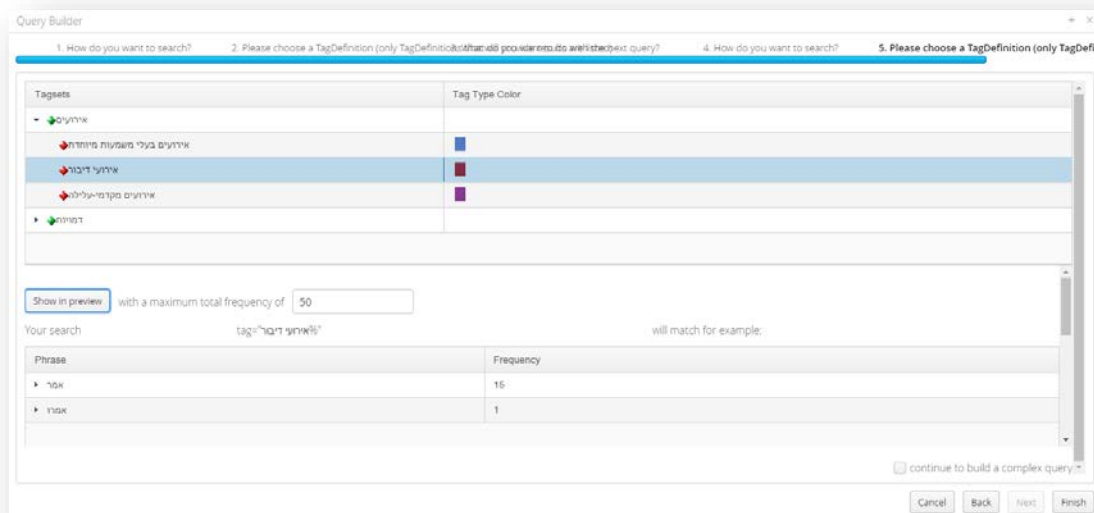
בין הממצאים בשלבים השונים של השאילתא – זאת,

משום שדי לנו במקרה זה בחפיפה חלקית בין התגים

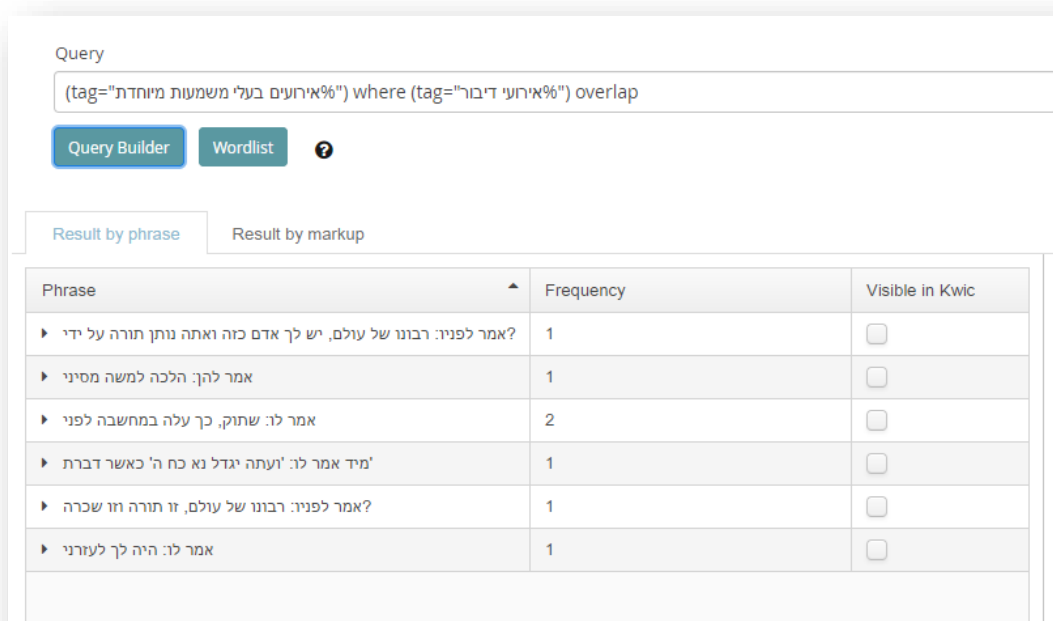
הרצויים. לאחר מכן נלחץ על כפתור Next בפינה

הימנית-תחתונה של החלון.

17.8. בחלון הבא נבחר שוב בתצורת חיפוש by Tag ונלחץ על Next. כעת נסמן את התג הרצוי – הפעם, 'אירועי דיבור' – ונלחץ על כפתור Show in preview. יושם אל לב: התוצאות הראשוניות המתקבלות כעת כוללות את כל אירועי הדיבור שתויגו בטקסט, ולא רק את אלו שסומנו גם כ'אירועים בעלי חשיבות מיוחדת'; זאת, משום שהאינטגרציה בין שני שלבי השאילתא עדיין לא נעשתה, ובינתיים כל שלב מוגדר באופן עצמאי.



17.9. הפעם נותיר את התיבה שליד continue to build a complex query, משום שברצוננו לסיים את הגדרת השאילתא ולעבור לקבלת הממצאים המשולבים. לחיצה על כפתור Finish תסיים את המהלך, ותוביל אותנו אל התוצאות הבאות:



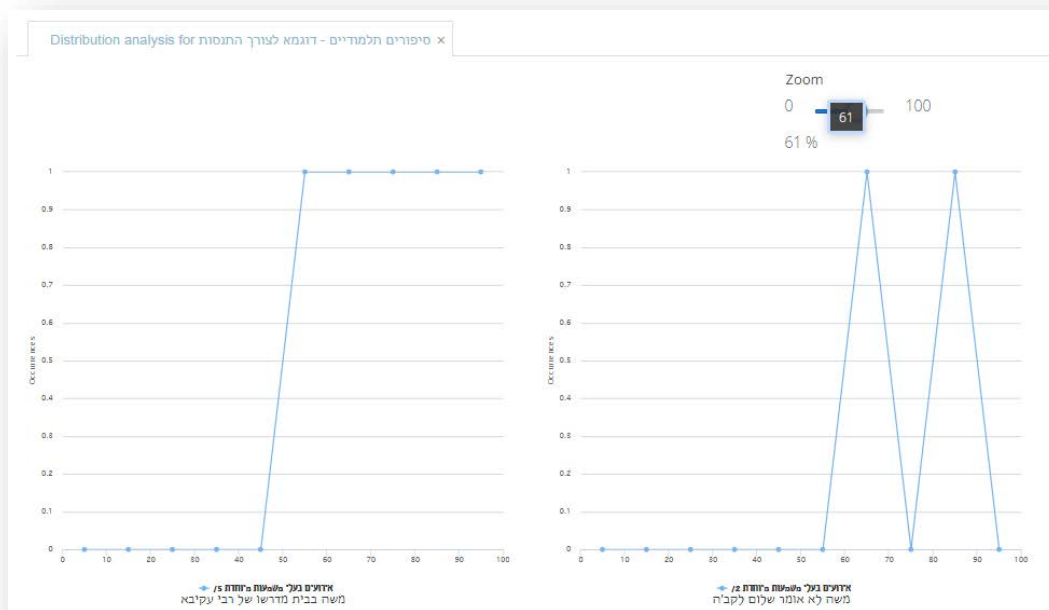
17.10. די במבט שטחי כדי לראות שכל התוצאות שהתקבלו (7 תוצאות בסך הכל) הן, אכן, 'אירועים בעלי משמעות מיוחדת' (בהתאם למידע שסופק קודם על ידינו למערכת בשלב התיג) – אך רק כאלה שיש בהם פעולות דיבור (המילה 'אמר' מופיעה בכל התוצאות). כדאי להתעכב רגע גם על ניסוח השאילתא בשורת הפקודה בחלקו העליון של המסך, המבהירה בלשון קצרה ובהירה למדי כיצד התקבלו תוצאות אלה דווקא:

אם ננסח את הפקודה במילים פשוטות נמצא כי השאילתא שבנינו מבקשת מהמערכת להציג כל מה שסומן בתג 'אירועים בעלי משמעות מיוחדת', אך רק כאשר (where) הם חופפים (overlap) לתג 'אירועי דיבור'.

17.11. בהתאם לדוגמה שנבחרה, ולאופרציונליזציה החישובית של שאלת המחקר, מה שמעניין אותנו כעת אינו דווקא לשון הממצאים – כל אחד משבעת הממצאים שונה בלשונו מהאחרים, שהרי המכנה המשותף להם אינו בטקסט, אינו ב- data, אלא בהפשטה המושגית שלנו, כלומר, ב- Markup collection, או ב- metadata; מה שמעניין אותנו, לעומת זאת, הוא **מקומם** של אירועים כאלה בקורפוס, ותהא אשר תהא לשונם. לכן נעבור מתצוגה בלשונית ה- Result by phrase לתצוגה בלשונית ה- Result by Markup. שבע התוצאות השונות 'מתורגמות' אוטומטית לתוצאה אחת ('אירועים בעלי משמעות מיוחדת') שלה שבע היקרויות:

Tag	Type Definition	Frequency	Visible in Kwic
▶	אירועים בעלי משמעות מיוחדת /	7	<input type="checkbox"/>

17.12. ושוב, מה שמעסיק אותנו אינו הממצא כשלעצמו, אלא מקומו בקורפוס; לשם כך נסמן את שורת הממצא המבוקשת, ונלחץ על כפתור גרף התפוצה שבפינה השמאלית-תחתונה של המסך. לחיצה זו תוביל אותנו ישירות למסך ה- Visualyzer, שם נקבל את שני הגרפים הבאים:



17.13. הגרפים מלמדים אותנו, למשל, כי בשני הסיפורים שבקורפוס אין אירועי-דיבור בעלי משמעות מיוחדת

בחציו הראשון של כל סיפור; זהו, כשלעצמו, ממצא מעניין מבחינת המבנה וה'מקצב' העלילתי שלהם (נקודה מעניינת במיוחד בהתחשב בתפקידה של השתיקה בכל אחד מהם). עוד הם מלמדים אותנו, כי בעוד סיפור 'משה בבית מדרשו של רבי עקיבא' (חמישה 'אירועים בעלי משמעות מיוחדת' נתונים שם בדחיסות כלשהי), סיפור 'משה לא אומר שלום לקב"ה' מגלם מהלך מינורי יותר מבחינה דרמטית.

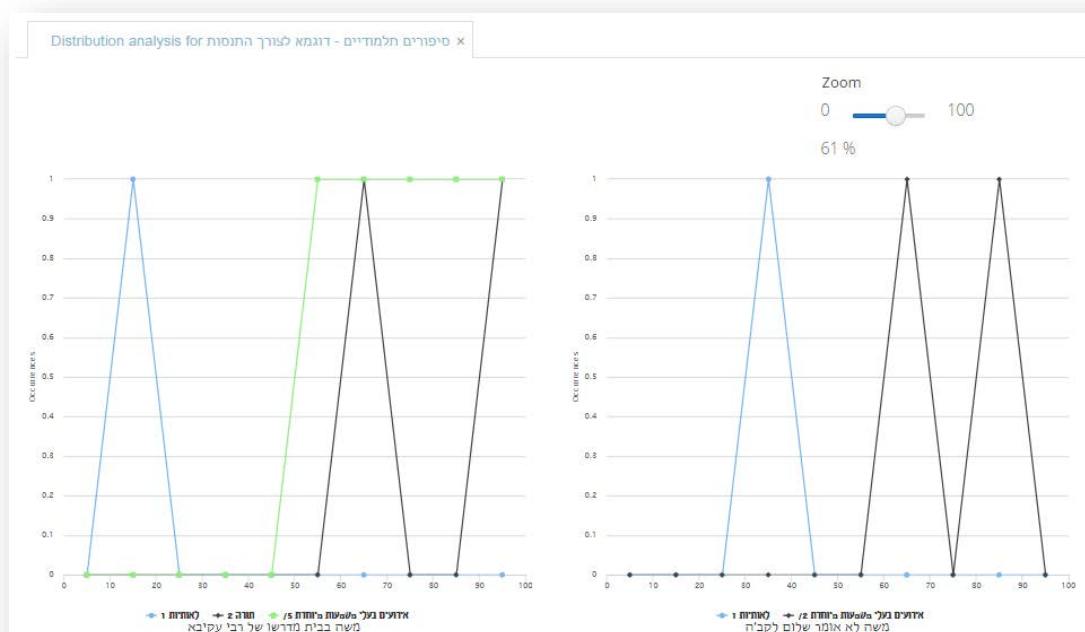
17.14. ואולם, כדי להשלים את התרגיל, נבקש לבחון את הממצאים העלילתיים בהשוואה לשני אלמנטים

לקסיקאליים 'אובייקטיביים' פשוטים. מעניין במיוחד יהיה להשוות את המבנה העלילתי המצטייר בגרפים, להופעת היבטים 'חומריים' יותר של מתן תורה – למשל, של המילה 'אותיות' ושל המילה 'תורה' עצמה. לשם כך נחזור אל מסך ה-Analyzer, ונבקש מהמערכת להפיק רשימת מילים (Wordlist). ממצאי השאילתא יעלמו מן המסך (אף על פי שהם ממשיכים להיות מתועדים בגרפים במסך ה-Visualyzer), ובמקומם תופיע רשימת המילים המרכיבה את הקורפוס. יושם אל לב: לקבלת רשימת המילים, עלינו לבחור בתצוגה תחת לשונית Result by phrase, שהרי מילים – בהגדרתן – אינן מסומנות מיסודן על פי תג (by markup).

17.15. כעת נסמן את המילה 'לאותיות' ונלחץ על כפתור גרף התפוצה; נציץ רגע במסך ה- Visualyzer, ושוב

נחזור למסך ה- Analyzer. נסמן את המילה 'תורה', נלחץ על כפתור גרף התפוצה, וכעת נתבונן בגרפים

המחודשים במסך ה- Visualyzer:



17.16. הגרפים שלפנינו מספרים סיפור מעניין ביותר: אם בנקודת המוצא של שני הסיפורים נראה כי ההיבטים

החומריים של התורה ניצבים במוקד – כתיבת התורה ואותיותיה, או קשירת הכתרים על גביהן (כפי שמעידה נוכחות המילה 'לאותיות' – פעם אחת בלבד בכל סיפור, בתחילתו) – מתברר כי המבנים העלילתיים מתרחקים מאוד מהיבטים אלה, ועוברים לצדדים מטאפוריים או סימבוליים יותר של המתרחש. ועם זאת, המילה 'תורה' נעדרת לגמרי מסיפור 'משה לא אומר שלום לקב"ה', בעוד היא מתועדת בזיקה הדוקה יחסית לשיאה המתמשך של עלילת הסיפור 'משה בבית מדרשו של הקב"ה'; חשיבותה שם אינה דווקא כשל אובייקט חומרי האמור לרדת עוד מעט קט לארץ, אלא כמכלול סמלי טעון, רווי משמעויות.

17.17. הנה כי כן, שילוב ממצאים מסוגים שונים – בתצוגה אחת, מרובת שכבות ופנים – העלה אפשרויות

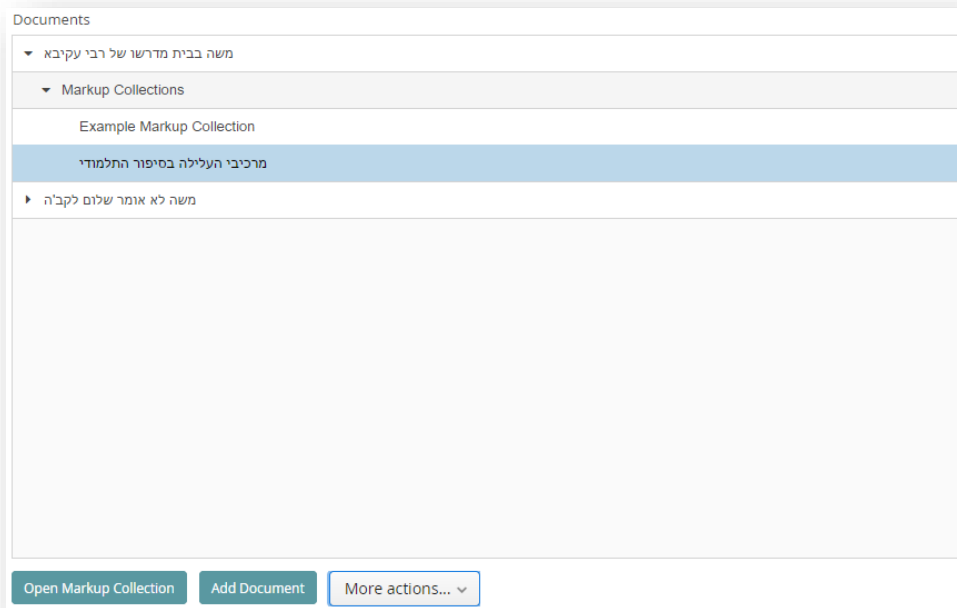
מאלפות בניתוח הקורפוס, ועתיד לעשות זאת ככל שנבקש. תועלתה של קריאה חישובית קפדנית כזו מתבררת אפוא לא רק במחקר ממושך של קורפוסים גדולים, אלא גם בדגימה קטנה ושטחית של קורפוס זעיר. אמנם ממצאים אלה, כמוכן, אילמים מצד עצמם; כדי להפיק מהם מחקר של ממש עלינו לדרב אותם ולתרגם אותם לתובנות משמעותיות. CATMA לא עושה עבורנו את כל העבודה, אך היא מספקת נקודות תצפית חדשות לגמרי להתבוננות במושא עיונו.

18.1. מערכת CATMA שומרת בחשבון המשתמש את הטקסטים שהזין לתוכה, כמו גם את ספריות התיוג ואת אסופות התגים (אמנם לפי שעה היא אינה שומרת שאילתות). כיאה לעבודה במערכת שנועדה לתמוך בפרוייקטים ממושכים ומסועפים, פעמים רבות נבקש לעיין שנית בטקסט שכבר תוייג בעבר – אם כדי להתבונן בשכבת התיוג ולנתח אותה, ואם כדי לשנות את התיוג או להוסיף על גביו שכבות תיוג חדשות. בבואנו לעשות זאת עומדות בפנינו שתי אפשרויות:

18.2. ראשית, נוכל לפתוח את הטקסט תוך התעלמות גמורה משלכיו הקודמים של הניתוח; לשם כך נסמן את שם הטקסט המבוקש במסך ה- Repository Manager, ונלחץ על כפתור Open Document. הטקסט ייטען למסך ה- Tagger ללא כל תיוג. בשיטה זו המשך שלבי הניתוח אינם שונים במאום מטיפול בטקסט חדש לגמרי, כפי שתואר בהרחבה בפרקי המדריך לעיל.

18.3. שנית – וזו, כמובן, האפשרות המומלצת – נוכל לפתוח אותו עם אסופת התיוג שכבר נוצרה, בין אם בכוונתנו להתייחס אליה, בין אם לאו (במילים אחרות, פתיחת הטקסט עם אסופת התיוג שלו אינה מייצרת בהכרח 'רעש' מיותר, משום שנוכל תמיד להחליט 'להסתיר' את השכבות הישנות). לשם כך נלחץ על המשולש הקטן שמשמאל לשם הטקסט המבוקש במסך ה- Repository Manager. כעת נלחץ על המשולש הקטן שמימין למילים Markup Collections, המופיעות מתחת לשם הטקסט.

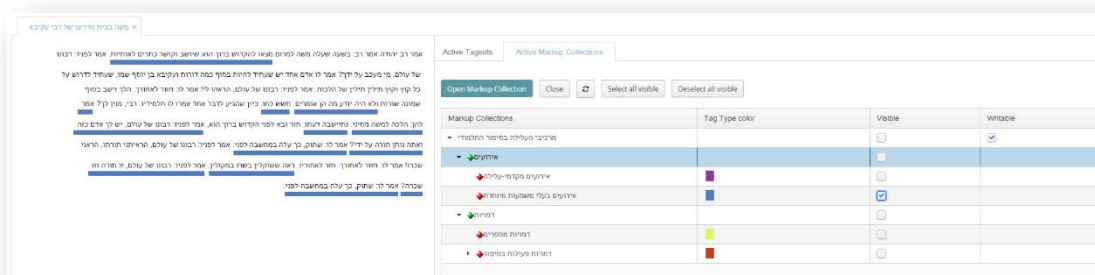
18.4. לחיצה זו תחשוף את אסופת התיוג שנוצרו זה מכבר לטקסט הנדון. משנסמן את האסופה הרלוונטית, כפתור Open Document ישנה את שמו ל- Open Markup Collection.



18.5. לחיצה על כפתור Open Markup Collection תוביל אותנו למסך ה- Tagger. הטקסט שבחרנו יופיע

במקום המיועד לו בצידו השמאלי של המסך; בצד ימין נראה את הלשונית Active Markup Collection פתוחה, ובתוכה טבלה המפרטת את תכני אסופת התיוג שנבחרה, בחלוקה לתגים ולתגי-משנה (כדי לראות את תגי-המשנה נלחץ על המשולש הקטן שמימין לכל תג).

18.6. סימון ✓ בריבועים הרלוונטיים שבעמודת ה- Visible תאפשר לנו לבחור אילו תגים ותגי-משנה יופיעו על גבי הטקסט המתוייג בצידו השמאלי של המסך. נוכל לבחור את כל הריבועים, את חלקם, או לא אחד מהם.



18.7. על פי ברירת המחדל של המערכת, אסופת התגים שדרכה פתחנו את הטקסט נועדה לשמש אותנו גם בתיוג

הקרוב. לפיכך הריבוע שבעמודת ה- Writable מסומן ב- ✓ מראש.

18.8. עם זאת, במצבים מסוימים נעדיף לנהל את התיוג החדש באסופה נפרדת. לשם כך נסיר את סימן ה- $\sqrt{\quad}$, ובמקום זאת נפתח אסופה חדשה בעזרת כפתור Open Markup Collection, וכפי שהוסבר בהרחבה לעיל.

18.9. כך או כך, כאשר נבקש לעבור לשלב התיוג-בפועל ונלחץ על לשונית Active Tagesets, נופתע אולי לגלות כי הטבלה שתחת לשונית זו ריקה. הסיבה לכך, כרגיל, היא עקרון הפרדת השכבות המאפיין את המערכת. גם אם ביקשנו לראות את הטקסט המתוייג, אין פירוש הדבר בהכרח שברצוננו להמשיך לתייג עם אותה מערכת תגים. לשם כך נלחץ על כפתור Open Tagset, ונטען את מערכות התגים הרצויות לנו, כפי שהוסבר לעיל. התיוג החדש יצטרף לקודמו – אם הותרנו את אסופת התיוג הישנה פעילה – או יישמר באסופה חדשה, בהתאם להחלטתנו.

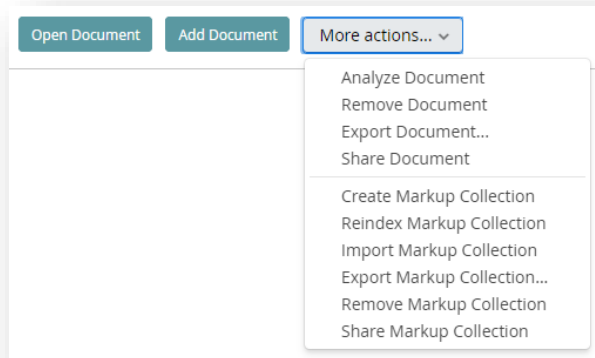
18.10. יושם אל לב: בחירת התגים המופיעים על גבי הטקסט במסך ה- Tagger אינה משפיעה כהוא זה על אפשרויות השימוש בהם והניתוח שלהם במסך ה- Analyzer (אליו נגיע או לאחר לחיצה על כפתור Analyze document במסך ה- Tagger, או לאחר בחירה באפשרות דומה תחת כפתור More Actions... הנמצא תחת רשימות הטקסטים והקורפוסים במסך ה- Repository Manager). במילים אחרות, גם אם לצורך עיון בתיוג הישן 'הסתרנו' חלק מן התגים ומתגי-המשנה, נוכל לבנות שאילתות מורכבות הכוללות גם אותם – ובתנאי שאסופת התיוג הרלוונטית נטענה למערכת.

19. ייצוא נתונים וממצאים, שיתוף מידע

19.1. מערכת CATMA מאפשרת ייצוא נתונים וממצאים כמעט בכל שלבי העבודה – אם לצורך גיבוי, אם לצורך עיבוד נוסף של החומר בתוכנה אחרת, אם לצורך הכנת המחקר לפרסום, ואם לצורך הצגתו בפני עמיתים ושותפים. בפרק זה תוצגנה תחילה אפשרויות הייצוא הפשוטות, ולאחר מכן יוצגו האמצעים לשיתוף פעולה עמוק יותר בין חוקרים.

19.2. כך אפשר לייצא מידע בצורת קובץ שישמר במחשבו של המשתמש:

19.2.1. ייצוא טקסט. נסמן את הטקסט המבוקש בטבלת המסמכים שבמרכז מסך ה-Repository Manager,



נלחץ על כפתור More

Actions... שתחת הטבלה,

ושם נבחר באפשרות Export

Document. בחלון הקטן

שנפתח נבחר את סוג הקובץ,

ולאחר לחיצה על Save המחשב

יפיק קובץ של הטקסט הרלוונטי.

19.2.2. ייצוא אסופת-תיוג. בדרך דומה, נסמן את אסופת התיוג המבוקשת בטבלת הטקסטים, נלחץ על כפתור

More Actions... ,ושם נבחר באפשרות Export Markup Collection. שוב נתבקש לבחור את סוג

הקובץ (והפעם, אם מדובר בקובץ המכיל גם את הטקסט או רק את שכבת התיוג) וללחוץ על Save. יושם

אל לב כי הקובץ מתקבל בתצורת XML, העלולה להיראות לגמרי בלתי-קריאה למשתמש חסר רקע.

19.2.3. ייצוא ספריית-תגים. נסמן את

ספריית התגים המבוקשת בטבלת

ספריית התגים בחלקו השמאלי-

תחתון של מסך ה-Repository Manager, ולאחר מכן נלחץ על כפתור Export Tag Type

Library. יושם אל לב כי הקובץ מתקבל בתצורת XML, העלולה להיראות לגמרי בלתי-קריאה למשתמש

חסר רקע.



19.2.4. ייצוא רשימת תוצאות. במסך ה-Analyzer, תחת רשימת התוצאות שבחלקו

השמאלי של המסך, נמצא כפתור קטן עם סמל תוכנת Excel. לחיצה עליו תגרום

למחשב להפיק קובץ גליון נתונים הכולל בתוכו את כל מה שברשימת התוצאות.

19.2.5. **ייצוא תוצאות בהקשר (Key Word In Context).** בדרך דומה, במסך ה-Analyzer, תחת טבלת

ה-KWIC שבחלקו הימני של המסך, נמצא כפתור קטן עם סמל תוכנת Excel. לחיצה עליו תגרום

למחשב להפיק קובץ גליון נתונים הכולל בתוכו את כל מה שבטבלת ה-KWIC.

19.2.6. **ייצוא גרפים.** במסך ה-Visualizer, מערכת CATMA אינה מספקת כפתור ייעודי, אלא מעודדת

אותנו דווקא להשתמש בצילומי מסך פשוטים (צילומי מסך מופקים בדרך כלל באמצעות לחיצה על כפתור

Ctrl ועל כפתור Print-Screen בו-זמנית. כפתור זה נמצא בדרך כלל באזור הימני-עליון של המקלדת);

צילומי המסך הללו ניתנים לעיבוד נוסף בכל תוכנה גרפית רגילה המותקנת במחשבו של המשתמש

(ברבות מתוכנות אלה ניתן לחתוך קטע מן התמונה – וכך אפשר, למשל, לשלב גרף במאמר).

19.3. מלבד טכניקות הייצוא הפשוטות הללו, אחד מיתרונותיה המובהקים של CATMA טמון בהיותה מערכת

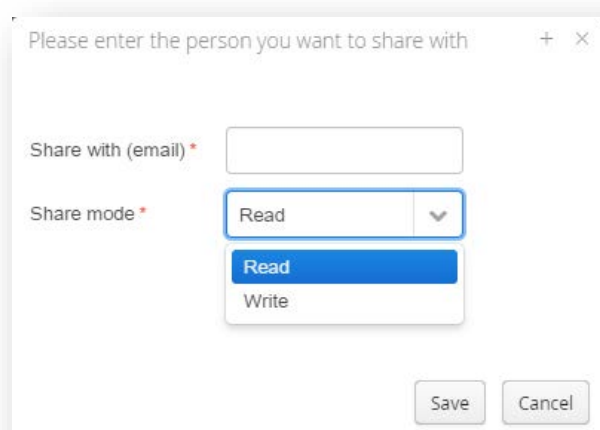
מקוונת, המאפשרת שיתוף מידע ברוב שלבי העבודה. כך, למשל, קבוצת חוקרים העובדת על פרוייקט משותף

תוכל להזין למערכת טקסטים, מערכות תיוג ואסופות תיוג, וכך לנהל את הניתוח הטקסטואלי באותה סביבה

דיגיטלית. חברי הקבוצה יוכלו לראות איש את תיוגיו של רעהו, להוסיף שכבות תיוג חדשות, ועוד.

19.4. כדי לשתף מסמכים נפתח את מסך ה-Repository Manager, נסמן את הטקסט המבוקש ברשימת הטקסטים,

נלחץ על כפתור More actions... הממוקם תחת הרשימה ואז נבחר באפשרות Share document. בחלון



שייפתח נקליד את כתובת הדוא"ל של מי
שעמו אנו מעוניינים לשתף את המידע,
ונגדיר את הרשאת השימוש שלו במידע:
האם יהיה באפשרותו לכתוב (write) –
כלומר, לשנות אלמנטים במידע המשותף
– או רק לקרוא (read), כלומר, להיחשף
למה שעשינו, ותו לא.

19.5. לחיצה על כפתור Save תגרום להוספת הטקסט שבחרנו לחשבון CATMA של

השותף. אם חשבוננו פעיל באותו זמן במחשב אחר, יהיה עליו ללחוץ על כפתור העדכון המובנה



במערכת על מנת שהמידע החדש יוצג לפניו. במסך ה- Repository Manager כפתור זה נמצא בפינה הימנית-עליונה.

19.6. באותה דרך בדיוק נוכל לשתף כמעט כל מידע אחר שהופק במערכת:

19.6.1. **שיתוף קורפוס.** במסך ה- Repository Manager נסמן את הקורפוס הרצוי ברשימת הקורפוסים, נלחץ

על כפתור More Actions... הממוקם תחת הרשימה, ונבחר באפשרות Share Corpus.

19.6.2. **שיתוף ספריית תגים.** במסך ה- Repository Manager נסמן את ספריית התגים הרצויה ברשימת

ספריות התגים שבחלקו השמאלי-תחתון של המסך. נלחץ על כפתור More Actions... הממוקם לידה,

ושם נבחר באפשרות Share Tag Type Library.

19.6.3. **שיתוף אסופת-תיוג.** במסך ה- Repository Manager נסמן את אסופת התיוג (Markup

Collection) הרצויה המופיעה תחת הטקסט הרצוי ברשימת הטקסטים, נלחץ על כפתור More

Actions... הממוקם תחת הרשימה, ונבחר באפשרות Share Markup Collection. אם מדובר

בטקסט שעדיין לא שותף עם הנמען, אסופת התיוג תעבור אליו אוטומטית יחד עם הטקסט הרלוונטי.

19.7. לשם הדגמה, הבה נניח שברצוננו לשתף פעולה עם חוקר העוסק גם הוא בסיפורים תלמודיים, אך מפרספקטיבה

שונה: הוא עמל על מיפוי הצורות הז'אנריות העיקריות בסיפורים אלה. לשם כך בנה אותו חוקר מערכת תגים

פשוטה בשם 'צורות ז'אנריות עיקריות בסיפורי חז"ל', ובה שלושה תגים: 'סיפור דרשני' (כלומר, סיפור שבמוקדו

עומדות דמויות מקראיות, והוא מפתח, משנה או מרחיב את העלילה המקראית), 'מעשה חכמים' (כלומר, סיפור

שבמוקדו עומדות דמויות מעולמם של חז"ל), ו'משל'. המוטיבציה לשיתוף פעולה, במקרה זה, הדדית: ברצוננו

לבחון את היחסים בין ניתוח הנראטיבים התלמודיים הקצרים, למיפוי הסוגתי שלהם. כיצד נבנה את הפרוייקט

המשותף במערכת CATMA?

19.7.1. תחילה על שני הצדדים להחליט על מידת השיתופיות – כלומר, אם ברצונם לעבוד על (א) טקסט/קורפוס משותף ואסופת תיוג משותפת; (ב) טקסט/קורפוס משותף ואסופות תיוג נפרדות; (ג) טקסטים/קורפוסים נפרדים ואסופות-תיוג נפרדות.

19.7.2. שנית, עליהם לתכנן את המחקר תוך מודעות לכך ששיתוף המידע אינו סינכרוני באופן מלא (כפי שמציעות תוכנות שיתופיות מסוימות דוגמת Google Docs, שבהן כל צעד של אחד השותפים נראה בזמן אמת על מסך המחשב של עמיתיו) אלא תמיד מתנהל בהשהיה כלשהי: כל צד תורם את חלקו ואז מעדכן מחדש את רעהו, וכן להיפך.

19.7.3. כך, בחזרה לדוגמא: במקרה שלנו החליטו החוקרים לעבוד על טקסט משותף, אך להעמיד על גביו שתי שכבות של אסופות תיוג נפרדות.

19.7.4. לשם כך הוחלט שהחוקר המתעניין בניתוח מבנים נראטיביים (להלן, חוקר מס' 1) ישתף את חוקר הצורות הז'אנריות (חוקר מס' 2) בטקסט 'משה לא אומר שלום לקב"ה', ובאסופת התיוג שבנה עבורו. לשם כך הוא פועל במסך ה- Repository Manager על פי ההנחיות שתוארו לעיל.

19.7.5. חוקר מס' 2 פתח במסך ה- Tagger בחשבוננו את הטקסט החדש ואת אסופת התיוג, וטען לתוכו מערכת תגים חדשה בשם 'צורות ז'אנריות עיקריות בסיפורי חז"ל', וכן אסופת תיוג מתאימה.

19.7.6. לאחר מכן תייג חוקר מס' 2 את הטקסט. במקרה זה התיוג היה פשוט: הטקסט כולו תויג בתג אחד רציף – 'סיפור דרשני'.

19.7.7. כעת משתף חזרה חוקר מס' 2 את חוקר מס' 1 בתובנותיו – כלומר, משתף אותו באסופת התיוג החדשה.

19.7.8. חוקר מס' 1, שלגמרי במקרה עובד באותו זמן על אותו טקסט בחשבוננו, מרענן את אסופת התיוג בעזרת כפתור הרענון במסך ה- Repository Manager. כעת הוא יכול לראות במסך ה- Tagger את שתי אסופות התיוג, שלו ושל חברו, בדיוק כפי שחברו רואה את שתי אסופות התיוג במחשבו שלו.

19.7.9. כל אחד מן החוקרים יכול להחליט אילו תגים (או אף אסופות תיוג שלמות) יוצגו על המסך, ואילו יוסתרו

באופן זמני, באמצעות סימוני \sqrt במקומות המתאימים בעמודת ה- Visible שתחת הלשונית Active

Markup Collections.

19.7.10. עתה מבקשים שני החוקרים להשוות את ממצאיהם גם ביחס לטקסט השני. חוקר מס' 1 משתף עם חוקר

מס' 2 את הטקסט 'משה בבית מדרשו של רבי עקיבא', ואת אסופת התיוג שבנה עבורו; חוקר מס' 2 מרענן

במסך ה- Repository Manager את אסופות התיוג, ופותח במסך ה- Tagger את הטקסט החדש ואת

אסופת התיוג שלו.

19.7.11. הוא טוען לתוכו אסופת תיוג ומערכת תגים, אך הפעם מתקשה בתיוג: האם במקור סיפור זה עומד משה

– ואזי יש לסווגו כ'סיפור דרשני' – או שמא רבי עקיבא, ואזי יש לסווגו כ'מעשה חכמים'? לאחר התלבטות

קצרה, מחליט חוקר מס' 2 להשאיר את ההתלבטות על כנה, ולתייג את הסיפור כולו פעמיים: כלומר, גם

כסיפור דרשני, וגם כמעשה חכמים.

19.7.12. בתום התיוג, משתף חוקר מס' 2 את חוקר מס' 1 באסופת התיוג שבנה; חוקר מס' 1 מרענן מחדש את

חשבונו, וטוען למסך ה- Tagger את שכבת המידע החדשה. שני החוקרים יכולים כעת לדון בעולה מצירוף

שני המחקרים יחד.

19.7.13. כך למשל נראה סיפור 'משה בבית מדרשו של רבי עקיבא' לאחר שתוייג בידי שני החוקרים; ראו כמה

שכבות! ועוד היד נטויה:

The screenshot displays the Tagger application interface. On the left, a text document is open, showing a passage from the Talmud with multiple layers of colored annotations (tags) applied to different parts of the text. The annotations are color-coded and some are visible through others, indicating multiple layers. On the right, there is a sidebar with two main sections: 'Active Tagsets' and 'Active Markup Collections'. The 'Active Markup Collections' section is currently selected and shows a list of collections with checkboxes for 'Visible' and 'Writable'. The collections listed include 'מרכיבי העלילה בסיפור התלמודי', 'אירועים', 'דמויות', 'מילות לאגדות בסיפורי תנ"ך', 'מילות לאגדות עיקריות בסיפורי תנ"ך', 'שמות', 'פסוקי דרשה', and 'מילים חכמיות'. Each collection has a 'Tag Type color' column and checkboxes for 'Visible' and 'Writable'. The 'Visible' column has checkboxes for each collection, and the 'Writable' column has checkboxes for each collection. The 'Tag Type color' column shows a small colored square next to each collection name. The 'Active Tagsets' section is currently empty.

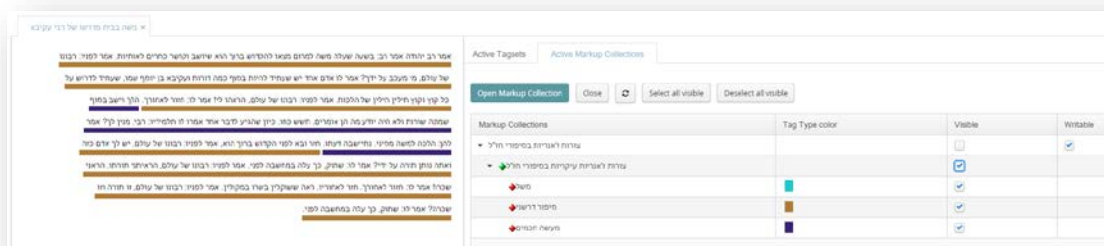
19.7.14. יושם אל לב: שני החוקרים צריכים לא רק להקפיד לשמור על סדר פעולות נכון, אלא גם על **תיאום מדויק**

של **הרשאות השימוש באסופות התיוג**. אין פירוש הדבר שכל שיתוף חייב להיעשות בתצורת Read ולא בתצורת Write, שכן לעיתים נמצא ערך מיוחד במעורבות השותפים באותה אסופת תיוג; ומבחינה טכנית, כל חוקר שקיבל הרשאה לכתוב יכול לבחור שלא לממש אותה – השליטה בכך נתונה בידי כל חוקר במסך ה- Tagger, תחת לשונית Active Markup Collections, בעמודות 'Visible' ו- 'Writable' (לשונית זו היא אולי האזור החשוב ביותר במערכת כשעובדים על מחקר מרובה משתתפים). לעומת זאת, חוקר שקיבל הרשאה לקריאה בלבד, לא יוכל לכתוב בלי קבלת הרשאה מחודשת. החלטה בעניין, מכל מקום, נתונה בידי החוקרים, במסגרת קביעת מידת השיתוף ביניהם.

19.7.15. כך למשל בדוגמא שהוצגה לעיל, חוקר מס' 1 סבר כי תיוג הסיפור כולו בשתי צורות ז'אנריות מופרז, וכי במקום זאת מוטב היה להבחין בין חלקים שבהם הדמות הדומיננטית בסיפור היא משה רבנו, לבין אותו חלק שבו הדמות הדומיננטית היא דווקא רבי עקיבא, בעוד משה צופה במתרחש באופן פאסיבי; יתכן גם שסבר כי כאשר משה יושב בבית המדרש של רבי עקיבא, הוא הופך כביכול לדמות מעולמם של חכמים, ולו לרגע.

19.7.16. מסיבות השמורות עמו, חוקר מס' 2 הואיל לאפשר לחוקר מס' 1 לשנות את התיוג בהתאם למסקנותיו. לשם כך, על חוקר מס' 2 לשותף עם חוקר מס' 1 גם את ספריית התגים (ואת זאת יעשה במסך ה- Repository Manager), וכן להרשות לו פעילות חופשית בכל השכבות (כלומר, שיתוף בתצורת Write – הן ביחס לאסופת התיוג, הן ביחס לספריית התגים).

19.7.17. לאחר התיוג המחודש נדרש חוקר מס' 1 לשותף את חוקר מס' 2 באסופת התיוג המעודכנת; חוקר מס' 2 ירענן את אסופות התיוג שלו במסך ה- epository Manager (לעתים יהיה עליו לצאת מהמערכת ולהיכנס אליה שוב, אם אינו מצליח לטעון את האסופה מחדש תוך כדי עבודה במערכת), ויקבל את אסופת התיוג שיצר בעצמו – אך בארגון חדש, והפעם כדעת חברו-שותפו:



19.7.18. בפני החוקרים עומדת גם אפשרות נוספת: הם יכולים להחליט שלא לגעת איש באסופת התיוג של רעהו, ובמקום זאת לייצר אסופת תיוג חדשה, מכוונת-שיתופיות, שאותה ורק אותה ינהלו יחד, תוך השוואתה לאסופות שגובשו באופן פרטי בידי כל אחד מהשותפים. לשם כך יהא על אחד מהם להגדיר אסופה חדשה בלחיצה על Open Markup Collection, ובחירה ב- Create Markup Collection. לאחר מכן יהיה עליו לשתף אותה עם חברו דרך מסך ה- Repository Manager, ולהמשיך לעבוד יחד בדרכים שתוארו לעיל.

19.8. ככל שהעבודה המשותפת הולכת ומסתעפת, כן מתגלה מעלתו של עקרון הפרדת השכבות ב- CATMA. האפשרות להעמיד זו על גב זו אינסוף שכבות תיוג, ביניהן כאלה התומכות זו בזו או סותרות זו את זו, כמו גם האפשרות להציג או להסתיר כל שכבה או כל פרט משכבה בכל שלב, מסייעות להעמיד מחקר טקסטואלי קבוצתי קפדני ומעמיק במיוחד.

20.1. להשלמת התמונה, נחזור בקצרה על העקרונות המנחים לעבודה במערכת CATMA – והפעם, בלי המימד הטכני המוציא אותם מן הכוח אל הפועל. עקרונות אלה ממצים את היסודות הקונספטואליים של המערכת, ויש בהם כדי להבהיר את היתרונות הייחודיים הגלומים בה:

20.1.1. **נגישות ותאימות.** מערכת CATMA אינה דורשת ידע בתכנות או במדעי המחשב. היא פונה ישירות לחוקרים מנוסים בפרקטיקות השגרתיות ביותר של מדעי הרוח, בהציעה להם סביבת עבודה רבת עוצמה, מותאמת לצרכיהם ומדברת בשפתם.

20.1.2. **שילוב קריאה קרובה (close reading) וקריאה רחוקה (distant reading).** דימוי רווח של פרקטיקות פרשניות המזוהות עם מדעי הרוח הדיגיטליים מתמקד ביכולתו של המחשב להציע טיפול סטטיסטי שטחי בכמויות מידע עצומות (Big-data), ולפיכך מצטייר לעיתים כלא-רלוונטי מזווית מבטו של חוקר האמון על קריאה אינטנסיבית בפרגמנטים קצרים. מצד שני, חוקרים האמונים על ניתוח כמויות מידע עצומות נוטים לעיתים להעצים את ערכם של ממצאים אמפיריים 'אובייקטיביים' המבטלים כביכול תובנות סובייקטיביות מקומיות. מערכת CATMA חותרת תחת דימויים אלה בהציעה סביבה אינטגרטיבית אחת לניתוח אוטומטי אמפירי ולתיוג פרשני יצירתי וחופשי המשלימים זה את זה ומאתגרים זה את זה חליפות.

20.1.3. **תנועה חופשית בין data ל- metadata ובין צורות ניתוח שונות.** תיאורטית, חמשת המסכים של המערכת יכולים להיות פעילים במקביל במהלך העבודה. פירוש הדבר הוא שהחוקר יכול לשמור על תנועת רצוא-ושוב בין (א) ניהול הטקסטים והקורפוסים העומדים לניתוח, לצד ספריות התיוג הפעילות, במסך ה- Repository Manager, ל- (ב) קריאת טקסטים וסיווג תופעותיהם במסך ה- Tagger, ל- (ג) הגדרה מחודשת ואפיון מחודש של הקטגוריות הפרשניות עימן ניגש אל הטקסט או הקורפוס במסך ה- Tag Type Manager, ל- (ד) ניתוח כמותי של כל אלמנט או תרכובת אלמנטים בטקסט/בקורפוס עצמו (data) או בקריאתו המתועדת על גביו (metadata), או בשניהם גם יחד, במסך ה- Analyzer, ל- (ה) ייצוג ויזואלי של ממצאים כמותיים מכל סוג במסך ה- Visualyzer. יתר על כן: האבחנה בין המסכים השונים אינה חדה מדי; חלק מהם מאפשרים עריכת פעולות מסוימות המזוהות כביכול עם מסך 'אחר' בלי

לעבור אליו. כך למשל מסך ה- Tagger מאפשר הגדרה מחודשת של מערכות התיוג לפרטיהן, ומסך ה- Analyzer מאפשר תיוג אוטומטי-למחצה של ממצאי החיפוש.

20.1.4. **רב-שכבתיות, יצירתיות וגמישות.** מערכת CATMA מפרידה לא רק בין הטקסט והקורפוס (ה- data) לשכבת התיוג הפרשני המוסבת עליהם (metadata), אלא גם בין אינספור רבדים של שכבה זו עצמה. החוקר (או צוות החוקרים) יכול לנהל אינספור אסופות-תיוג (Markup collections) על גבי הטקסט, ולהחליט בכל רגע נתון אילו מרכיבים מתוכם יוצגו, ואילו יוסתרו. אין גבול לשילובים האפשריים ולמשחק הפתוח-לגמרי בין תפיסות שונות של מושא המחקר.

20.2. סיכם את הדברים יאן כריסטוף מייסטר :

מזווית מבט קונספטואלית ומתודולוגית מהלך העבודה הזה הוא, כמובן, פרקטיקה מחקרית רגילה, רק כזו המשתמשת בכלי חישובי. אך האם היא מובנת מאליה? לא ככל שמדובר בהיבט החישובי. משום שהמודל הקונספטואלי ההרמנויטי והפילולוגי המונח בתשתית CATMA איננו עוד שרשרת קלט-פלט מכאנית. למעשה, הוא מחזורי ובלתי-מוגבל. ניתוח, תיוג ופרשנות צועדים כאן יד ביד, וכל אחד יכול להיכנס, לצאת, לסיים או להתחיל מחדש את רצף הפעולות הללו בכל נקודה שיבחר... אך יש כאן יותר מאשר מחזוריות: בשונה מכלי תיוג אחרים, CATMA מאפשרת לעשות בדיוק את מה שנראטולוגים ופילולוגים בני-אנוש עושים כל הזמן: לחלוק, להתווכח, לסתור אחד את השני, לסתור אפילו את עצמם. מתייג אחד עשוי למצוא סיבה טובה לסמן פיסקה בטקסט כאנלפסיס [= חזרה לאחור בזמן הסיפור. א.מ.], בעוד מתייג שני סבור שאותה פיסקה עצמה ראויה להיחשב לייצוג של דיבור ומחשבה פנימיים. CATMA עוצבה כך שתקבל ותשמור את שתי האפשרויות, והיא תמשיך לעשות זאת עד מספר אינסופי של קוראים.¹⁰

20.3. ומעל לכל, אמנם, נדרשת **סבלנות**. מחקר חישובי מן הסוג שתואר במדריך זה אינו מציע קסמים מהירים המושגים בלחיצת כפתור, אלא שיטת עבודה יסודית ואפילו תובענית. תיוג רציני אורך זמן; קטגוריות פרשניות משתנות עם התקדמות המחקר; לא תמיד מובן בנקל מה יש לחפש; ולעיתים הממצאים דווקא ברורים, בעוד השאלה שעליה הם עונים חבויה בערפל. סוף סוף, החזרה המתמדת אל הטקסט המתוייג מעמידה תמונת-ראי בלתי מתפשרת נגד עיני החוקר; אך אחריתה – מי ישורנה.

¹⁰ Jan Christoph Meister, 'Toward a Computational Narratology', in: M. Agosti and F. Tomasi (eds.), Collaborative ¹⁰ Research Practices and Shared Infrastructures for Humanities Computing, Padua: CLEUP, 2014, p.29.

21.1. להשלמת המידע על אודות תכונות המערכת והשימוש בהן, כדאי להיעזר בחומרי ההסבר וההדרכה המגוונים

[שבאתר המערכת](#) (הן אלו שתחת לשונית Functionality, הן אלו שתחת לשונית Ducomentation). חומרים

אלה מתעדכנים מעת לעת, וכוללים גם [מערכי-תרגול](#) המבוססים על טקסטים שהוזנו למערכת למטרה זו. בין

היתר נמצא שם [מדריך מקיף באנגלית](#), הנוקט גישה טכנית ומינימליסטית יותר בהשוואה למדריך העברי. תועלת

מובטחת גם למשתמשים שיעזרו במידע המתקבל בלחיצה על כפתור סימן-השאלה המלווה את מסכי המערכת.

21.2. להעמקה נוספת בהבנת ההגיון הפרדיגמטי והקונספטואלי של CATMA, ראו מאמרו של אבי המערכת: Jan

Christoph Meister, 'Toward a Computational Narratology', in: M. Agosti and F. Tomasi (eds.),

Collaborative Research Practices and Shared Infrastructures for Humanities Computing,

Padua: CLEUP, 2014, pp. 17-36. המאמר מתייחס אמנם גם לכמה היבטים באוטומציה של תהליך העבודה,

שאנם רלוונטיים לפי שעה למחקר טקסטים עבריים, אך במידה רבה אפשר לראותו כמעין 'מניפסט' של המערכת.

21.3. לדוגמת מחקר שנערך באמצעות CATMA ראו מאמרה של Evelyn Gius, ['Narration and](#)

[Escalation: An Empirical Study of Conflict Narratives'](#), Diegesis 5.1 (2016), pp. 4-25.

לדוגמא שונה במקצת ראו מאמר המשלב את CATMA עם כלים חישוביים נוספים: Stuart Dunn and

Mareike Schumacher, ['Explaining Events to Computers: Critical Quantification,](#)

[Multiplicity and Narratives in Cultural Heritage'](#), Digital Humanities Quarterly 10.3

(2016).

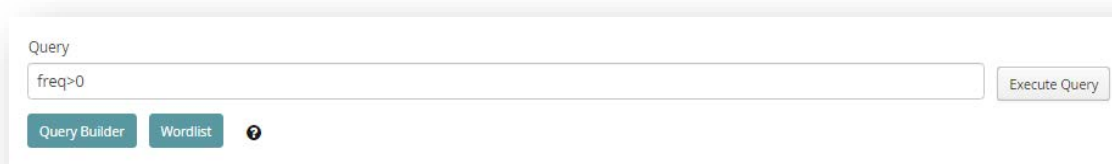
21.4. לדיון עקרוני בשאלת תפקידה של פרספקטיבה כמותית מדידה במחקר ספרותי (ובמידה רבה, במחקר טקסטואלי

אופייני למדעי הרוח בכלל) ראו מאמרו של Franco Morreti, ['Operationalizing: Or, the Function](#)

[of Measurement in Modern Literary Theory'](#), Stanford Literary Lab Pamphlet no.6

(December 2013).

22.5. כבר נאמר לעיל כי הדרך הנוחה ביותר ללימוד שפת השאלות של CATMA – עבור אלה מהקוראים שאינם מסתפקים בבונה-השאלות הידיוותי-למשתמש – היא לעקוב אחר ניסוח השאלות המופיע במסך ה-Analyzer בגמר יצירת השאלות. המלצה זו מקבלת משנה תוקף בהתחשב בקשיים הנגזרים משילוב שפות שונות, כפי שיתוארו להלן.



22.6. כתיבה ידנית של שאלות לטקסטים בעברית עשויה להיות מאתגרת מעט מבחינה טכנית, שכן בעוד מילות הקבע של הפקודות, על שלל הסימנים המוסכמים שלהן, נכתבות באנגלית ונקראות משמאל לימין, הרי שהטקסט המתבקש על ידינו נכתב עברית ונקרא מימין לשמאל. שילוב שתי שפות, שני כיווני קריאה, סימנים מוסכמים ואף מספרים, הוא פתח לבעיות רבות בייצוג השאלות בחלון המיועד לכך (בעיות אלה מוכרות לכל מי שמשלב שפות שונות במסמכים, בדואר אלקטרוני ובכל כיוצא בהם). המערכת יודעת אמנם 'לקרוא' גם שפות הכתובות מימין לשמאל, אך מזווית מבטו של המשתמש שורת הפקודה עשויה להיות מבלבלת למדי: סימני הפיסוק נראים כאילו אינם במקומם, סדר המילים משובש, והתחביר הכולל של הפקודה מעורפל. לבעיה זו אין פתרון פשוט מלבד הקפדה על כתיבת כל אחד ממרכיבי השאלות לפי הסדר: סדר הפעולות המתבקשות מן המערכת הוא משמאל לימין, בעוד סדר הקריאה הפנימי של המילים והתווים בעברית הוא מימין לשמאל. הודעה על תקלה בשלב זה עשויה להיות אינדיקציה לשיבוש כיוון הקריאה, ומחייבת כתיבה מחודשת של הפקודה.

22.7. עם זאת, לא פעם נמצא כי כתיבה ידנית של שאלות נוחה יותר, ומאפשרת ניסוח מהיר של קומבינציות מורכבות למדי. כמו כן נמצא כי הבנת שפת החיפוש של המערכת מאפשרת תיקון, שינוי ושיפור מהיר של שאלות שנכתבה במקורה על ידי בונה השאלות; לעתים מתברר אפילו שהשאלות המסוימת שלה אנו זקוקים פשוט אינה ניתנת לניסוח מודרך בבונה-השאלות. מטרת הנספח, אפוא, היא להציג כמה פקודות בסיסיות, שימושיות במיוחד, המדגימות מקצת ממגוון הפעולות האפשריות בשפה זו, וממחישות את המבנה התחבירי שלה.

22.8.1. כדי לחפש מילה מסוימת נקליד את המילה בין מרכאות כפולות. למשל: "אדם".

22.8.2. כדי לחפש צירוף מסוים נקליד אותו באותה דרך בין מרכאות כפולות. למשל: "רבנו של עולם".

22.8.3. סימני פיסוק נחשבים במערכת ליחידות-משמעות עצמאיות, ולכן החיפוש "אמר לו שתוק" לא יוביל לתוצאה כלשהי, משום שלא כך הוזן הטקסט למערכת; חיפוש מדויק יותר, התואם את האופן שבו הוזן הטקסט למערכת, יפתור את הבעיה. כך: "אמר לו: שתוק". עבור מחקרים מסוימים די בעובדה זו כדי לעודד הסרה מוקדמת של סימני הפיסוק, חלקם או כולם, בשלב התקנת הטקסט; אולם מחקרים אחרים עשויים דווקא למצוא בכך תועלת.

22.9. חיפוש צירופים רווחים:

22.9.1. כאשר לא כל הפרטים ידועים, ובכל זאת אנו מעוניינים לחפש דפוסים רווחים שונים המבוססים על מידע מוקדם חלקי, נציב בתחילת הפקודה את מילת ההקדמה `reg=` (קיצור של `regular expression`), בתוספת המידע הרלוונטי:

22.9.1.1. כדי לחפש את כל הצירופים הכוללים אות ידועה אחת או יותר, נשתמש בנקודות במקום התווים שאינם ידועים – מספר הנקודות כמספר התווים. כך למשל הפקודה `reg="ה..ה"` בקורפוס הגדון כאן תניב את התוצאות 'הלכה', 'הראה' ו'הודה'; ואילו הפקודה `reg="...ה"` תניב תוצאות כגון 'ישבה' ו'תורה' (יושם אל לב כי המחשב קורא מילים בעברית כסדרן, מימין לשמאל, אך סימנים שאינם אותיות בעברית נקראים כבאנגלית, משמאל לימין, כך שהאות ה' בפקודה האחרונה מופיעה בסוף צירוף בן ארבעה תווים). אותה פקודה בדיוק תניב גם תוצאות כגון 'לא ה' וכן 'ו, ה' (שהרי סימני פיסוק ורווחים נחשבים אף הם לתווים).

22.9.1.2. כדי לחפש את כל הצירופים הכוללים אות ידועה אחת או יותר, ואפשרות לא-הכרחית לאות נוספת ואפילו לחזרה על אות זו, נשתמש בכוכבית לאחר האות המסופקת. כך למשל הפקודה

"*מש" reg= פירושה, בקירוב: אתר עבורנו את כל הצירופים המתחילים באות מ', שאחריה האות ש' מופיעה אפס פעמים או יותר. פקודה זו תניב את כל צירופי ה'מש' (בקורפוס שלנו צירופים אלה מופיעים תמיד בשם 'משה'), אך גם את כל הופעותיה של האות מ', שכלל אינה סמוכה לאות ש'. בקורפוס אחר אפשר שנקבל כך גם מילים כגון 'משש' ו'מששלחתי'.

22.9.1.3. בשפות אחרות, שבהן רצפי עיצורים זהים שכיחים יותר מבעברית, שימוש בכוכבית כפי שתואר בסעיף הקודם עשוי להניב תוצאות רבות. בעברית נוכל להסתפק בדרך כלל בסימן שאלה באותו מקום, והתוצאות תהיינה כמעט זהות. סימן השאלה יניב את כל הצירופים שבהם האות האחרונה שלפניו מופיעה פעם אחת או שאינה מופיעה כלל.

22.9.1.4. אם ברצוננו למצוא את כל הצירופים שבהם התו האחרון מופיע לפחות פעם אחת, נשתמש בסימן חיבור. כך למשל הפקודה "+ לפני" reg=, שבה סימן החיבור מוצב לאחר תו רווח, תניב בקורפוס שלנו אך ורק תוצאה אחת: 'לפני', ואילו הפקודה "? לפני" reg= תניב גם תוצאות כגון 'לפניו', 'לפני'.

22.9.1.5. אם ברצוננו לאתר את כל הצירופים בהם תו ידוע אחד או יותר מופיע לפני אחד מתוך כמה תווים ידועים, נציב את התווים האפשריים בתוך סוגריים מרובעים. כך למשל הפקודה "[על]" reg= תניב הן את כל הצירופים של 'שע' (כמו במילים 'שעה', 'שעתיד' ו'יהושע'), וכן את כל הצירופים של 'של' (כמו במילה 'שלום' או בצירוף 'של עולם').

22.9.1.6. כדי לשלול את הופעת אחד מתוך כמה תווים ידועים לאחר התו המבוקש, נוסיף לצד התווים שאינם רצויים, בתוך הסוגריים המרובעים, את הסימן הדיאקריטי הנראה כמעין V הפוך קטן, המקובל בשפות אחדות ומכונה בעברית 'גג'. כך למשל הפקודה "[ש^ע]" reg= תניב את כל הצירופים של ש' עם כל התווים למעט ע'.

22.9.1.7. כדי לאתר את כל הצירופים של תו ידוע אחד או יותר, עם סדרה אפשרית של אותיות הנתונות בטווח מסוים בתוך גבולות האלף-בית, נציב סימן חיסור בין גבולות הטווח בתוך הסוגריים המרובעים. כך למשל הפקודה "[א-ל]" reg= תניב צירופים כגון 'שב' (כמו במילה 'במחשבה'),

‘שי’ (כמו במילה ‘שיושב’) ו‘של’ (כמו במילה ‘שלום’), אך לא צירופים כגון ‘שע’ (כמו במילה ‘שעלה’) או ‘שת’ (כמו במילה ‘שתוק’), בהם התו הבא אחרי האות ש’ נמצא מחוץ לגבולות שנקבעו בסוגריים.

22.10. חיפוש מילים לא-זהות בעלות מאפיינים משותפים:

22.10.1. דרך אחרת לאיתור מילים לא-זהות בעלות מאפיינים משותפים היא להציב בתחילת הפקודה את מילת ההקדמה wild= (קיצור של wild card, תו חופשי), בתוספת המידע הרלוונטי. אפשרות זו מספקת פתרון משלים נוסף לבעיית השורש בשפה העברית, אך מלבד זאת יש לה יתרונות רבים אחרים. כך:

22.10.1.1.1. אם ברצוננו לאתר את כל המילים הלא-זהות שיש להם מכנה משותף כלשהו, נשתמש בסימן אחוזים, שתפקידו לייצג את החלק המשתנה של המילים. כך למשל הפקודה wild=% תניב את המילים ‘כתרים’ ו‘אומרים’.

22.10.1.1.2. אם ברצוננו להגביל את מספר התווים הלא-ידועים באותן מילים, נשתמש בקו תחתון – מספר הקווים כמספר התווים (בדומה לנקודות בפקודות מסוג reg). כך למשל הפקודה wild=_____ תניב את המילה ‘אומרים’, אך לא את המילה ‘כתרים’.

22.11. חיפוש מונחה דומות: כדי לחפש מילים שונות הדומות זו לזו במידה כלשהי, נשתמש בפקודת simil= (קיצור של similarity), שלאחריה המילה המבוקשת בין מרכאות, ושיעור הדומות באחוזים. כך למשל הפקודה simil=70% אמר תניב את התוצאות ‘אמר’, ‘ואמר’, ‘אמרו’. כפי שכבר נזכר לעיל, פקודה זו מספקת אף היא פתרון חלקי וזמני לסוגיית השורש בעברית.

22.12. חיפוש מונחה קרבה: כדי לאתר מילים מסוימות בתנאי שהן סמוכות למילים מסוימות אחרות, נציב את הסימן & בין שתי המילים הרלוונטיות (הנתונות כל אחת בתוך מרכאות), ואחריהן את טווח הקרבה במספרים. כך למשל הפקודה 5 משה & שלום תניב הופעה אחת בלבד של השם ‘משה’ – זו הסמוכה בטווח של חמש

מילים מהמילה 'שלוש' במשפט 'משה, אין שלוש בעירך?'. יושם אל לב כי איננו מחפשים כך את הצירוף, אלא משתמשים בו כאמצעי לסינון תוצאות מסוימות מכלל הופעות השם 'משה' בטקסט.

22.13. **חיפוש מונחה תדירות:** כדי לאתר מילים בהתאם למידת שכיחותן בטקסט או בקורפוס, נשתמש בפקודה freq (קיצור של frequency) שלאחריה הסימן המתמטי הרלוונטי, ושיעור התדירות המבוקש. נוכל לבקש תדירות מדויקת (=), תדירות קטנה מ- (<), גדולה מ- (>), קטן מ- ושווה ל- (<=), גדול מ- ושווה ל- (>=), ובטווח שבין ל- (-). כך למשל הפקודה `freq > 5` תניב תוצאות כגון 'של' (6 מופעים), 'אמר' (13 מופעים), 'לו' (6 מופעים), ואף סימני פיסוק המופיעים יותר מחמש פעמים בקורפוס הנדון כאן.

22.14. **חיפוש מונחה תג:** כדי לאתר את כל מה שתויג באופן מסוים על ידינו, נשתמש בפקודה `tag=`, ולאחריה שם התג המבוקש נתון בתוך מרכאות. כך למשל הפקודה `tag="מספר"` תניב את התוצאות 'רב יהודה אמר רב' ו'רבי יהושע בן לוי'. אם ברצוננו לאתר תת-תג מסוים, נצרף אותו לאחר התג שאליו הוא שייך, ונפריד ביניהם באמצעות לוכסן (המרכאות תישארנה בגבול החיצוני של התג). כך למשל הפקודה

`tag="דמויות פעילות בסיפור/דמות מקראית"`

תניב את התוצאה 'משה'.

אם ברצוננו להשאיר את שיוכו של תת-תג לתג אחר באופן פתוח, נוכל להציב סימן אחוזים במקום התג העליון יותר בהירארכיה (בדומה לשאלות מסוג wild). כך למשל הפקודה `tag="דמות אלוהית"` תניב את התוצאות 'רבנו של עולם', 'הקדוש ברוך הוא', 'ה', ואף 'משה' – משום שאחד המתייגים שמעשיו הודגמו במדריך זה סבר כי באחד הסיפורים עובר משה מטאמורפוזה.

22.15. **שאלות מורכבות:** כזכור, שאלות מורכבות עשויות מיותר משלב אחד. אם עד כה למדנו כיצד לנסח כל שלב בנפרד, כעת נלמד להגדיר את היחסים ביניהם.

22.15.1. **הוספת תוצאות.** אם ברצוננו לבנות רשימת תוצאות המורכבות מצירוף פשוט של יותר משלב חיפוש אחד – בלי שהתוצאות של שלב אחד משפיעות על אלו של השלבים האחרים – נקליד פסיק בין השלבים השונים. כך למשל הפקודה `tag="משה", tag="דמויות מספרים"` תניב את התוצאות 'משה', 'רבי יהושע בן לוי', 'רב יהודה אמר רב'.

22.15.2. **גריעת תוצאות.** אם ברצוננו לגרוע תוצאות מסוימות מרשימה שגובשה בשלב אחד או יותר, נשתמש בסימן מינוס בין השלבים הרלוונטיים. כמו כן לא נשכח לציין לאחר השלב השני את מידת ההתאמה בין ממצאי השלבים השונים. כך למשל הפקודה

overlap "אירועי דיבור" tag=" - "אירועים מקדמי-עלילה" tag=

תניב בתוצאותיה את כל האירועים מקדמי-העלילה שאינם חופפים לאירועי דיבור.

22.15.3. **זיקוק תוצאות.** אם ברצוננו לזקק את רשימת התוצאות, כלומר, להתנות את הכללת תוצאות מסוימות בתוצאותיה של פעולת חיפוש אחרת, נשתמש במילה where כשהיא נתונה בין השלב העיקרי לשלב המשמש כתנאי עבורו. כך למשל הפקודה

overlap "אירועי דיבור" tag= where tag="אירועים מקדמי-עלילה" tag=

תניב בתוצאותיה את כל האירועים מקדמי-העלילה החופפים לאירועי דיבור (אפשר לנסח פקודה זו גם באופן אחר – על ידי שימוש בסימן נקודה-פסיק במקום המילה where. במצב כזה נמנע מלהגדיר את מידת ההתאמה בין שלבי החיפוש).

22.15.4. כאמור, נספח זה מפרט את הפקודות העיקריות בלבד בשפת השאילתות של המערכת. להיכרות עם פקודות נוספות יש לעיין [במדריך האנגלי](#), שאף הוא אינו ממצה לגמרי את שלל האפשרויות. כך למשל כשנעיין בשאילתות המנוסחות על ידי המערכת באמצעות בונה-השאילתות, נמצא שימוש מרובה בסוגריים, כמו גם בשילוב פקודות רבות יחד. בסופו של דבר, אין תחליף להתנסות אינטנסיבית; ככל שנעמיק את נסיונו, כך נדע לנסח גם את השאילתות הסבוכות ביותר באופן ידני ומהיר.