



שיטוט ביער המידע

חדשות וחידושים מעולם הספרנות והמידענות

שלום לכולם וברוכים הבאים ל"שיטוט ביער המידע"!

הפעם הניזולטר מופיע בגרסת קיץ שונה מהרגיל, בהמשך להרצאה המרתקת של ד"ר גל יעבץ מהמחלקה למדעי המידע בבר אילן: איסוף ואיתור מידע באוקיינוס הביג דאטה ללא תכנות (No-Code). ההרצאה, שהתקיימה בזום ב-5/7/22, היא הראשונה מתוך סדרת המיטאפים "מומנטום: מידע בתנועה" – יוזמה משותפת של המחלקה למדעי מידע באוניברסיטת בר אילן עם הוועדה המתמדת של מנהלי הספריות האוניברסיטאיות.

כולנו מייצרים מידע בכל רגע נתון. כשאנחנו מחפשים ב-Google, כשאנחנו עורכים קניות בסופר, כשאנחנו מאזינים למוזיקה, מידע ונתונים על ההתנהגויות שלנו נאספים וניתן להפיק מהם תובנות משמעותיות. בהרצאה הוצגו מספר כלים מרכזיים ושימושיים שיכולים לסייע לנו, כמידענים ואנשים סקרנים, להגיע למסקנות מתוך נתונים שונים בדרך קלה ואפקטיבית. נעבור על כלים אלו כאן בהרחבה: חיפוש שאלות, מגמות וטרנדים בהתבסס על מסדי הנתונים של Google, חיפוש דאטה ונתונים ממשלתיים מ-Dataset, ניתוח טקסטים ואימוג'ים מ-Twitter, ניתוח מוזיקה מ-Spotify, וגם – איך לייצר של רשת מאמרים אקדמיים דרך Connected Papers.

כל הכלים הם חינמיים !

קריאה נעימה ושימושית,

עמליה

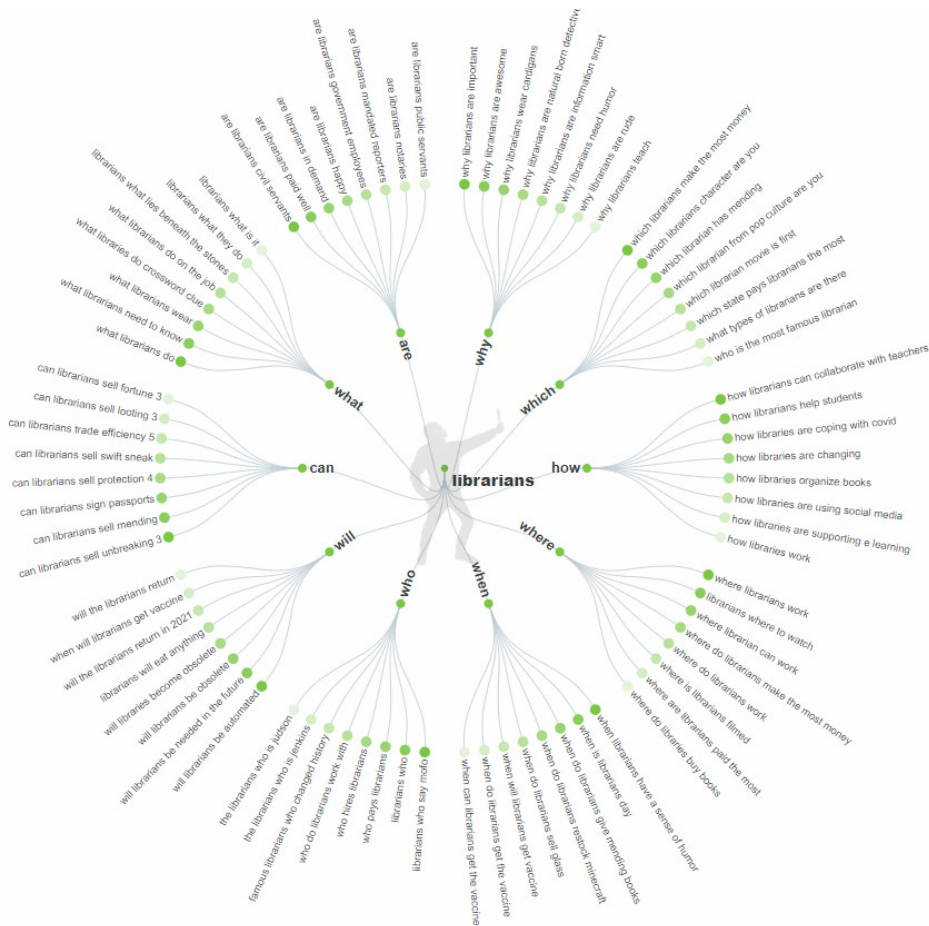
שאלות, מגמות וטרנדים דרך מסדי התונים של Google

בהרצאה הוצגו שלושה כלים המבוססים על מסדי התונים של Google:

Google Trends-ו Answer the public, Trendy.

Answer the public – כלי המאפשר לראות מה אנשים שואלים. הרבה מהחיפושים בגוגל הם לא רק מילה או נושא מסוים, אלא חיפושים הרבה יותר פרסונליים כמו "מהן שעות הפעילות של הספרייה?" "היכן אוכלים באוניברסיטת תל אביב?". וכך, הכלי מאפשר לחפש מילה או ביטוי בכל השפות, לסנן לפי מקום גאוגרפי ולראות מה נשאל בהקשר שלו במנוע החיפוש. למשל חיפוש המילה **librarians** מראה לנו מה אנשים שואלים על ספרנים.

הנתונים מוצגים באמצעות ויזואליזציה מרשימה של גלגל ענק, חלוקה לפי מילות שאלה ובהמשך לפי מילות יחס, חלוקה אלפביתית והקשרים נפוצים.



חיפוש Librarians -ב Answer the Public

בין התוצאות אפשר למצוא תהיות פילוסופיות והתחבטויות מקצועיות, לצד שאלות על סדרת טלוויזיה ושלל פנינים נוספות :

Librarians what is it? ?Why librarians are important Librarians what they do?
?Librarian of farmer? Why librarians wear cardigans

אפשר לשנות את הוויזואליזציה לתצוגה טבלאית ולהוריד את המידע לקובץ CSV.

זהו כלי המהווה קו ישיר לדעת הקהל ולכן מועיל לכל העוסקים בקידום אתרים. למשל, אפשר להיעזר בו כדי לבחור כותרת אטרקטיבית לפוסט בבלוג או כדי להבין על מה כדאי לשים את הדגש בבניית אתר אינטרנט. מעניין גם לבחון דרכו הבדלים תרבותיים בין מדינה למדינה, דרך ניסוחי השאלות השונים העולים ממנו.

*יש הגבלה על מספר החיפושים החינמיים שאפשר לעשות ביום.

Trendy – כלי נוסף המבוסס על מסד הנתונים של google, מאפשר לראות חיפושים פופולאריים מהיומיים האחרונים, על מה אנשים מדברים הכי הרבה בעולם או באזור מסוים. מדובר למעשה בנושאים שחלה עלייה בכמות החיפושים שלהם, דרך טובה להבין על מה מדובר בגוגל רגע נתון ואולי להחליף צפייה בחדשות

הנתונים מוצגים באמצעות בועות כחולות, לחיצה על כל אחת מהבועות תוביל לכתבה רלוונטית לנושא שחיפשנו. אם הבועה גדולה במיוחד סימן שגם העלייה בכמות החיפושים הייתה משמעותית.



*תוצאות החיפוש ב- Trendy מה- 19.7.2022

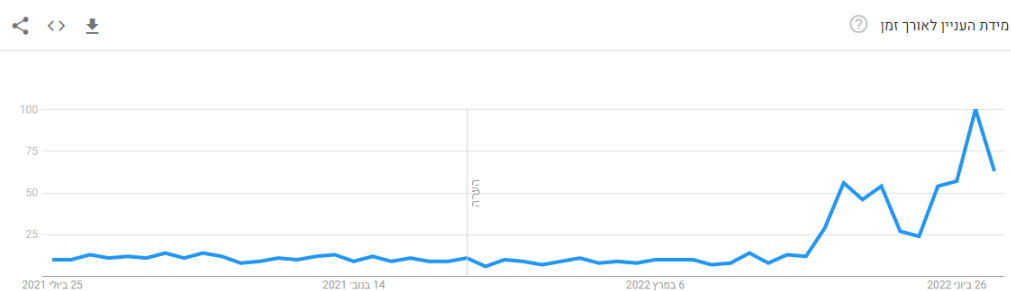
Google trends – זהו הכלי המוכר מהבין שלושת כלי גוגל שהוזכרו בהרצאה (ואפילו הוזכר בניוזלטר הראשון שלנו). למרות שרבים מכירים אותו, לא רבים מודעים לשלל היכולות המתקדמות שלו. גם פה אפשר כמובן לחפש בשפות שונות ובאזורים שונים

בעולם ובטווחי זמן מגוונים: שעה / יום / שנה / 5 שנים / זמן מותאם אישית ועוד... ולהשוות בין ערכים שונים.

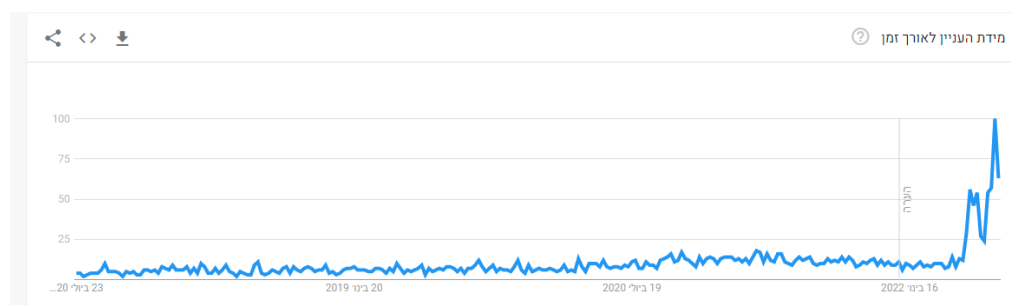
לדוגמא: חיפוש **ספריית סוראסקי** בשנה האחרונה. אפשר לראות שיש עלייה בחיפושים בתקופות מסוימות כמו תחילת סמסטר ואפילו עליה מסוימת בקיץ וירידה בתקופות אחרות כמו חופשת הסמסטר.



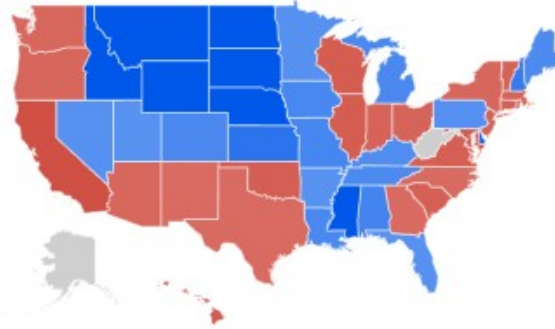
דוגמה נוספת: עלייה בפופולאריות של המילה פיסטוק בשנה האחרונה



אם מסתכלים על טווח של חמש שנים זה אפילו מובהק עוד יותר...האם מדובר במגמה חשובה או טרנד חולף? בשביל זה צריך להמשיך לעקוב.



בנוסף, ניתן לערוך השוואות שונות באמצעות מפות: איפה מתעניינים בגביע העולם בכדורגל? היכן עומדת לפרוץ שפעת? את מי חיפשו במדינות ארצות הברית השונות בשבוע האחרון יותר - נתניהו או לפיד?



[Page Views](#) - כלי משלים לשלושת כלי גוגל שנסקרו לעיל. הוא שייך לוויקיפדיה ונותן מידע על צפייה בערכים ספציפיים לאורך זמן, תוך אפשרות לעריכת השוואות שונות. אפשר גם להתייחס לטווח תאריכים ספציפי ולהתייחס להבדלים לפי ימים או לפי חודשים (שימו לב, כשנכנסים לאתר הוא באופן אוטומטי מראה לנו השוואה בין המונחים Cat ו-Dog).

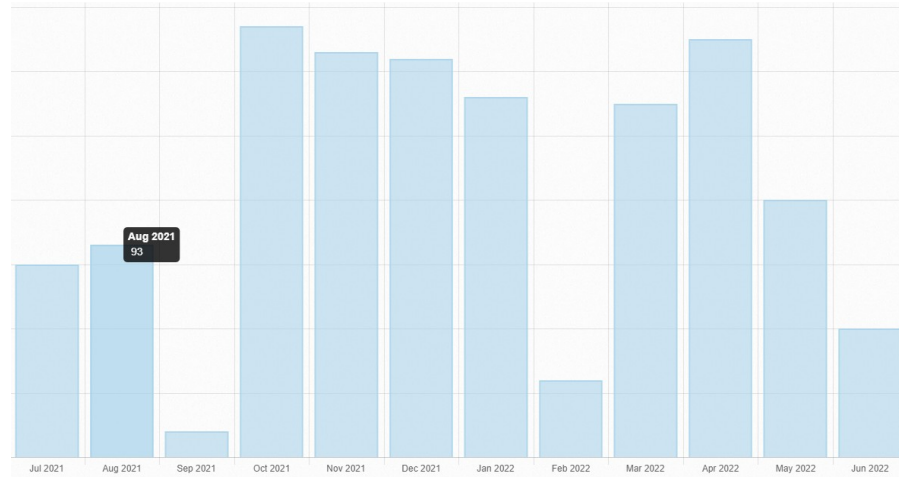
דרך נוספת להגיע לנתוני הצפייה היא מדפי הערכים עצמם. בוחרים ב-'מידע על הדף' תחת כלים באפשרויות המופיעות בצד ימין של הערך ולאחר מכן לוחצים על 'קישור לגרף סטטיסטיקת צפייה באתר wmf labs'



סטטיסטיקת צפיות

קישור לגרף סטטיסטיקת צפייה באתר wmf labs, עם אפשרויות נוספות

[הספרייה המרכזית ע"ש סוראסקי](#) גם כאן בחיפוש לפי חודשים, ניתן לראות שחודשי החיפוש המובילים במהלך השנה האחרונה היו אוקטובר ואפריל והפחות פופולריים היו ספטמבר ופברואר, בהתאם לזמני פתיחת שנת הלימודים וחופשות הסמסטר והקיץ.



זהו כלי פשוט מאוד לשימוש, המראה לנו עד כמה וויקיפדיה הפכה להיות כוח מוביל בכל הקשור למשאבי ידע.

מידע נוסף על טרנדים אפשר למצוא גם דרך [Twitter](#), עם הכניסה לאתר בצד ימין נמצא את כפתור Trends for you, דרכו נוכל לצפות בכל החיפושים המובילים באזור ברגע זה.

דאטה ונתונים

[Dataset Search](#) – כלי נוסף מבית גוגל, אך מעט שונה ממה שראינו עד עכשיו. זהו מעין מאגר מידע של נתונים המבוסס על מסדי נתונים של רשויות ממשלתיות וגורמים ציבוריים. בעזרתו ניתן להגיע למידע שיכול להיות רלוונטי לחוקרים ומדענים על משרדי ממשלה, שווקים ומדיניות.

הכלי מאפשר לחפש מידע מאוד ספציפי בעזרת שאילתות מורכבות, כמו – מה היו מותגי הקפה המובילים בספרד ב-2020? מי שינסה למצוא את התשובה דרך מנוע החיפוש הרגיל של Google ילך לאיבוד בתוך בליל של תוצאות לא בהכרח רלוונטיות. דרך Dataset אפשר לקבל את התשובה באמצעות חיפוש בסיסי ופשוט.

המאגר מאפשר סינונים כמו: צפייה רק בנתונים חנימיים ובחירת סוג המסמך (תמונה, טקסט, טבלה). אפשר לראות גם מי הוא הספק של הנתונים ולהוריד קבצי Excel.

בינתיים המאגר עובד פחות טוב בעברית מכיוון שאין מספיק הנגשה של מסמכים ונתונים רשמיים בארץ. ארגונים כמו 'שקוף', 'התנועה לחופש המידע' ו'ישראל דיגיטלית' פועלים לפרסום הנתונים, לצד אינדוקס מסוים מ-Data.gov, אבל המאמצים הללו עדיין אינם מספקים.

עוד על המאגר מתוך [גיקטיים](#) ו- [עולם המידע](#).

כלים מבוססי Twitter

[Vicinitas](#) – כלי לניתוח מידע מ-Twitter, שהיה עד לא מזמן אתר נישתי של עיתונאים. הוא קטן בהשוואה לרשתות חברתיות אחרות, אך ממשיך לצמוח ולגדול כל הזמן וניתן למצוא בו מידע ושיח שניתן להסיק מהם תובנות מעניינות. לכלי יש גרסה חינמית וגרסה בתשלום, הדרישה היחידה היא להיות מחוברים ל-Twitter. אפשר לחפש מידע על תגיות או מילות מפתח מובילות, על משתמשים או עוקבים מסוימים, למשל מי עוקב דמות ספציפית. ניתן גם לראות את כל הציטוטים בנושא מסוים (בגרסה החינמית עד 2000 ציוצים מעשרת הימים האחרונים). השימוש בכלי פשוט ואינטואיטיבי.

גם פה חיפשתי את המילה [פיסטוק](#) (נראה שעל סוראסקי אין מספיק ציוצים) וקיבלתי 533 תוצאות. ניתן לייבא את הנתונים לקובץ Excel ולראות את הציוץ המלא, מי המשתמש?, מתי?, מי עשה לו ריטוויט? ואפשרויות נוספות כמו לארגן את הציטוטים לפי מספר הלייקים.

בשל עומס הנתונים, ד"ר יעבץ ממליץ להיעזר בכלי נוסף ששמו [Voyant](#). זהו כלי לניתוח טקסט מעולם מדעי הרוח הדיגיטליים שתורגם לעברית על ידי ד"ר סיני רוסניק מאוניברסיטת חיפה, המאפשר לנתח כמויות גדולות מאוד של טקסט בצורה אוטומטית. דרכו ניתן לראות מילים נוספות החוזרות על עצמן בציוצים, לאתר ציוצים מובילים וליצור ניתוחי טקסט שונים.

[EmojiTracker](#) – כלי חביב העוסק בניטור אימוג'ים בזמן אמת ב-Twitter. דרכו אפשר לראות כל אימוג'י שעושים בו שימוש ברגע נתון. ככה זה נראה :

| | | | |
|--------------|---------------|--------------|--------------|
| 👍 3644001258 | ❤️ 1912312270 | 😬 1588627020 | 😬 1289798581 |
| 👍 438627112 | 😬 428610092 | 😬 427801502 | 🔥 424366520 |
| 👍 264557886 | ❤️ 261847424 | 🎵 257474636 | 📧 240465102 |
| 👍 189518523 | 👍 186371123 | 😬 183889470 | 😬 181428321 |
| 🎉 146630929 | 🐱 145414543 | 🌸 142474315 | 😬 141046006 |
| 📧 113834645 | 🌑 112773939 | ☀️ 112269323 | 😬 112034373 |
| 👍 95243886 | !! 90085535 | ⚡ 89114841 | 👍 87226631 |
| 🎵 61231624 | 🎵 58907843 | 👤 58675044 | 👍 58034549 |

[EmojiTracker](#)* בזמן אמת

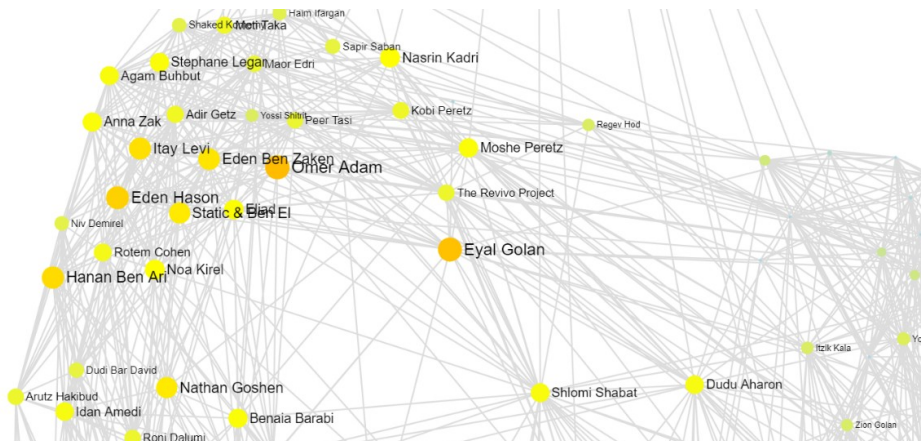
אם בוחרים את אחד האימוג'ים – אפשר לראות את הציוצים שנכתבים אתו כרגע, הסבר על המשמעות שלו ואת מיקומו במדרג הפופולאריות. זהו כלי קצת פחות שימושי

מבחינה מידענית, אבל הוא נותן אינדיקציה למה פופולארי ברגע זה ומהו הלוך הרוחות הציבורי.

מוזיקה ורשתות

ניתוח האזנה למוסיקה (spotify) – כל פעולה שלנו משאירה סימן בעולם הנתונים, וכך גם האזנה למוסיקה של האומן או הלהקה האהובים עלינו. גם במימד הזה נוכל להיעזר בכלים שונים להפקת תובנות, כמו בכלי שמתייחס להאזנה ב-spotify ועל סמך נתונים יוצר ניתוח רשתות של קשרים ומשרטט יחסי גומלין בין אומנים ולהקות בהתאם לאנשים המאזינים להם.

לדוגמא: חיפוש של עומר אדם מראה את הקרבה שלו לאומנים אחרים כמו עדן בן-זקן, עדן חסון וסטטיק ובן-אל בהתאם להעדפות הקהל. ככל שהיעיגול יותר גדול מדובר באומן שיש לו יותר מאזינים ברשת. באמצעות הוויזואליזציות הללו ניתן לזקק תובנות ומסקנות על העדפות האזנה, את מי שומעים ואת מי פחות שומעים וכדומה. ניתן להוריד את הניתוחים כקובץ JDF.

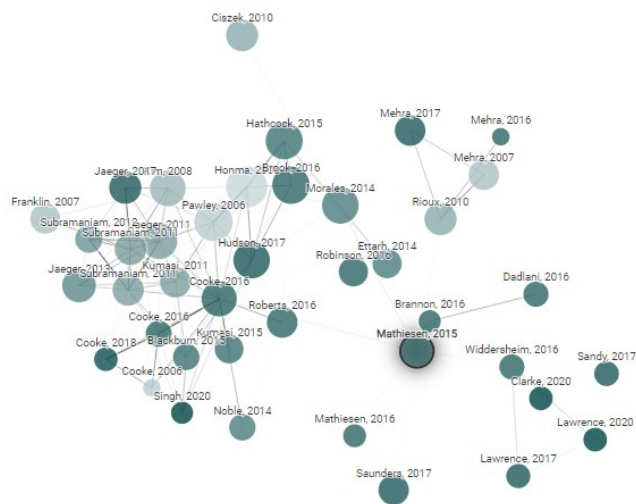


*ניתוח האזנה ל- [עומר אדם](#) דרך Spotify

רשתות באקדמיה

Connected Papers – ונסיים עם כלי המתאים לעולם האקדמיה. זהו פיתוח ישראלי של תלמידי מחקר המבוסס על הרעיון שכל פריט אקדמי הוא חלק מרשת: אפשר לראות את מי הוא מצטט, מי מצטט אותו ומה הקשרים בין כל הגורמים הללו. כל פריט הוא חלק ברשת של ידע אקדמי, כשאפשר גם לראות את מי כמעט ולא מצטטים והוא נשאר כעיגול מבודד. ד"ר יעבץ מלמד את הכלי יחד עם Scopus ו- Web of science וממליץ

עליו כדרך נוספת למצוא מאמרים דומים כשנקודת המוצא היא מאמר רלוונטי בודד.
את הנתונים המתקבלים ניתן להוריד כטבלה ולבצע עליהם ניתוחים סטטיסטיים



*דוגמא לרשת הקשרים של המאמר [Informational Justice: A Conceptual Framework for Social Justice in Library and Information Services](#) מתוך [Connected Papers](#)

עד כאן להפעם. ניפגש בניוזלטר הבא!