

# **הצגה מקובצת של תוצאות חיפוש על פי טקסטומיה ארגונית בשילוב מנשך LCC&K**

**יקי מרדכי, עפר דרורי, אריאל פרנק**

אחת הבעיות המרכזיות בתחום אחזור מידע מטעורה בשלב הבדיקה של רשימת התשיבות על ידי המשתמש. המאמר מציג את האחרון בסדרת מחקרים ש�示ם איתור המידע הרלוונטי ביותר להציג רשימת תוצאות חיפוש, כדי לסייע למשתמש לקבל החלטה נכונה; על אילו מסמכים תוצאות החיפוש להתקדם כדי לקבל מענה למשימת החיפוש שלו. אנו מציעים במאמר מודל חדש להציג תוצאות חיפוש TLS (ע"ץ טקסטומיה ארגונית בשילוב תקצרי מנשך LCC&K). המחקר בוצע עם 60 משתמשים, ונערכה בו השוואה בין השימוש במונחים תוצאות חיפוש החדש לעומת השימוש חיפוש ארגוני קיימ. מצאי הממחקר מראים כי בזמן החיפוש החדש קוצר יותר בהשוואה במנשך החיפוש המודד על מנת שמשתמש יוכל להגיע למסגרת המבוקש, ורוב (77 אחוז) משתמשי הניסוי העדיפו שימוש מונחי החיפוש היו כדוגמת מונשך TLS.

**מבוא**

עמ' מנסך K,LCC.

**פריטי מידע לגבי כל מסמך במסמך התשיבות**

באוניברסיטה העברית בירושלים נערכה סדרת מחקרים לאיתור המרכיבים העדיפים לשימוש בעניין המשמש, והשפעתם עליו בעת הצגת רשימת תוצאות חיפוש (Drori, 2000, Drori & Tamir, 2005; Drori & Alon, 2003 הממצאים הראו כי שימוש בקטגוריות המסמכים, שורת מהטקסט בהקשר החיפוש ומילוט מפתח, עדיף על פני הצגת אותו מידע ללא מרכיבים אלו. נמצא דומה הtgtala לגבי קטגוריות המסמך, והמודל שהומלץ יושם על ידי מנסך LCC&K. מודל K LCC&K מסתמך עדין רק על דירוג הרשימה, וכן ניתן לשפרה. כדי לקבל מידע לגבי מידת הדמיון או השוני בין המסמכים ברשימה, יש לסרוק את רשימת המסמכים כולה. המסמכים המשויכים לאזטה קטגוריה יכולים להופיע באופן מפוזר לאורך רשימת התשיבות הארוכה, וכך אין תמונה כללית טוביה של המבנה של הקטגוריות הרלוונטיות ושל פיזור המסמכים בתוכן.

במחקר שנערך על ידי אייבורי, יי' וגורונמאיר (Yu, Yaron, Gronemyer, 2004) עולה, כי משתמשים מעדיפים לראות פריטי מידע נוספים לגבי המסמכים, על פני היכולת לשלוט בצורות המין של הרשימה. לטענותם, משתמשים מרגשים שתהיליך מציאת התשובה יעל יותר בצורה משמעותית כשהם לא נכנסים למסמכים שאינם עונים על דרישותיהם. לפיכך, הצגת שורות נוספים לככל מסמך יכולת לסייע למשתמש להחליט אם עליו להיכנס למסמך או לא.

אחת הבעיות המרכזיות בתחום אחזור מידע מטעורה בשלב הבדיקה של רשימת התשיבות על ידי המשתמש. רוב מונחי החיפוש יכולים מציגים את התשיבות ברשימה רציפה, המדורגת על פי מידת הקربה של התשיבות לערכי השאילתת החיפוש. רמת הדירוג הנמוכה של המונחים, בשילוב עם הצגת הרשימה הרצופה, מקשים על המשתמש לאיתר את המידע הרלוונטי שchipsh. מלבד מספר התשיבות הגדול, שקרים לוודאי התקבלו כמענה לחיפוש, המסמכים ברשימה אינם מופיעים על פי סדר נושא כלשהו, ותהליך ניפוי התשיבות הופך לפיקר למסורבל. התהיליך קשה עוד יותר כאשר מדובר במסימות חיפוש מחקריות, שבהן יש יותר מסמך אחד רלוונטי ונדרשת הבנה נוספת של התוצאות ברשימה. המאמר מציג את האחרון בסדרת מחקרים ש�示ם איתור מרכיבי המידע הנחוצים ביותר מתוך רשימת תוצאות חיפוש, כדי לסייע למשתמש בקבלת החלטה אינטואטיבית ונוחה: אילו מתוך מסמכיו רשימת התוצאות כוללים בתוכם את התוצאה המבוקשת למשימת החיפוש.

**מחקר קודם**

הספרות המדעית הרחבה כוללת מחקרים שונים שנערכו לבחון השפעתם של מרכיבי מידע מלאה ואחרים על הצגת תוצאות חיפוש. מכיוון שמרכיבי המידע הנחוצים להציג תוצאות חיפוש בכל מסמך ומסמך ברשימה התוצאות נכר אונדרה בסדרת מחקרים מוקדמת, (Drori & Tamir, 2005; Drori, 2000) שהייתה את הבסיס לmanshak המחקרי -LCC&K Line in Context Categories and Key Words בתוספת הייחודית שמלהו שילוב ע"ץ קטגוריות של המסמכים ←

ד"ר עפר דרורי, offerd@gmail.com  
יקי מרדכי, mordechai@013.net  
ד"ר אריאל פרנק, ariel@cs.biu.ac.il, המחלקה למדעי המחשב, BIU

לשימוש, מעניק ביטחון רב יותר בתחום איתור התשובה, ואך מטעה פחות לעומת הממשק שבו לא היו קטגוריות (& Drotz, 2003; Alon, 2003).

## HFC (Hierarchical Faceted Categories)

חשוב להתמקדש בשימוש שנעשה בקטgorיה, בין אם זו נוצרה בתחום אוטומטי ובין אם הוגדרה מראש. קטgorיה מכוננת-תוךן הפקה נפוצה בשנים האחרונות, והיא מהווה חלק מודיע-על שיכול להתווסף למסקר. בראש קיימים מאגרים בנותאים כגון עיתונות ורפואה המכילים בנוסף למסמכים גם מידע-על שככלו בתוכו קטgorיות. התיאוג המצויר למסמכים מאפשר למעשה ליצור הצגה לוגית של המסמכים. טקסונומיה מכילה למעשה המאורגן לצורה המשקפת מרכיבים מרכזים של התחום שבו היא עוסקת. רישימת הקטgorיות נוצרת בדרך כלל באופן ידני, ותהליך שירוק המסמכים לקטgorיות יכול להתבצע לצורה ידנית או אוטומטי. לאחרונה מתפתחת ונחקרת הן בעולם האקדמי והן בעולם המסחרי גישת HFC. העיקרון המרכזי על פי גישה זו הוא שימוש במבנה מדריגי של קטgorיות המשicasות לכמה פנים. כל פן מהווה מעין קטgorיות-על או ממד לצורך בחינת אוסף מסמכים. לכל פן משicasות הקטgorיות הרלוונטיות לו. מבנה HFC מאפשר ניהול יווט במספר אופנים שונים במסגרת תהליכי ה"עיליה" ברישימת התשיבות, כאשר מידע-העל שכמודרג מסיע למשתמש (Hearst, 2006).

**ממשק מכוון קטgorיות**  
מחקר שנערך על ידי דומאים, קטרל וchan (Dumais, Cutrell, and Chen, 2001) הראה כי מנשכים שביהם מצחת רשימת תוצאות במנשך עיילים יותר ממנשכים שביהם מצחת רשימת תוצאות במנשך הבינו מקטgorיות נמצא עיל יותר גם ממנשך מקביל אשר מכיל חלק מהתוצאות כותרת המסמך את הקטgorיה אליה הוא משתייך. המשמשים העדיפים מנסכים מכוסטי קטgorיות, שהתגלו כמוירם יותר בזמנים אוחז במצב התשובה. מנסכים מכוסטי קטgorיות כוללים:

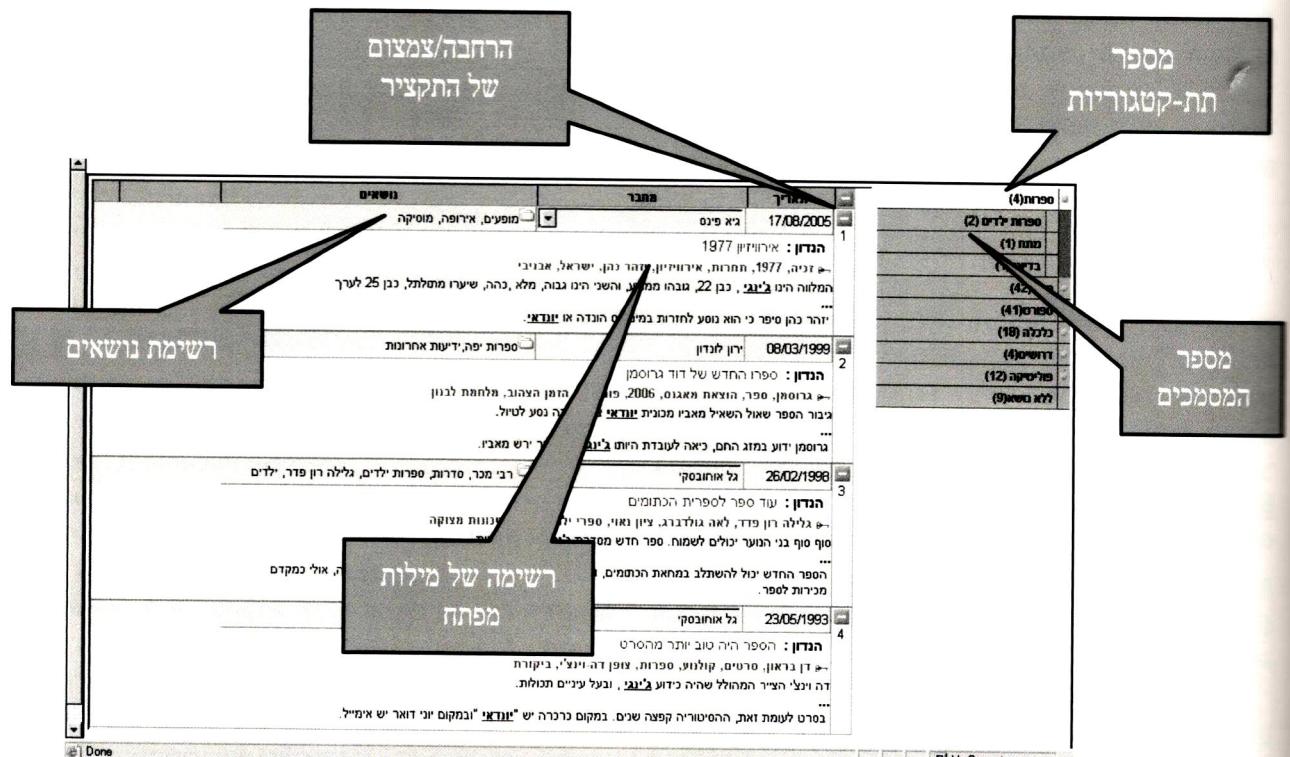
## INDEX

מנשך המחולל קטgorיות דינמיות על פי מילים וביטויים נפוצים בתחום המסמך. רכיב החיפוש עושה שימוש בספרית ריבוי התוכנה של מנוע גוגל. ההצגה נעשית בעץ בעל רמה אחת. כיוון שהאלגוריתם מכוסס על ניתוח סטטיסטי, אין התייחסות למשמעות המלים, ולכן אין ניתן שהקטgorיות יהיו משמעותיות. למעשה, נוצרים מצבים שביהם שמות הקטgorיות חסרים כל הקשר או שאינו מובנים למשתמש (Kaki, 2005).

## TLCC (Titles, Lines in Context and Categories)

במסגרת המחקר שהוביל לפיתוח ממשק LCC&K, נבדקה התועלות הנובעת מהוספת קטgorיות לרישימת תוצאות החיפוש. נמצא כי הרואו תועלת משמעותית למשתמשי ממשק מבוסס קטgorיות במגוון היבטים - ממשק מהיר יותר ונוח יותר

## איזור 1 - ממשק TTLS



$$tf_{Ci,q} = 0.5 + \frac{0.5 * f_{Ci}(q)}{\max_j f_{Cj}(q)}$$

$f_{Ci}(q)$  - מספר המטמכים המשיכים לקטגוריה  $Ci$  בשליפה  $q$   
חסכונות קטגוריה  $Ci$  במאגר  $D$ :

$$idf_{Ci} = \log \frac{|D|}{f_{Ci}(D)}$$

$|D|$  - מספר המטמכים במאגר  
 $f_{Ci}(D)$  - מספר המטמכים המשיכים לקטגוריה  $Ci$  במאגר  $D$   
 וצר-הכל עבור כל קטגוריה נקבע את משקלה על ידי מכפלה של שני הפרמטרים:

$$tf_{Ci,q} W_{Ci,q} = * idf_{Ci}$$

## אבטיפוס TTL&NISI

מטרת אבטיפוס TTL&NISI להראות כי המודל שהוצע אכן פותר את הבעיה העיקרית, לצורך בדיקת השיטה החדשה והערכתה, נבנה ניסי המשווה את מודול/מנשך TTL&NISI מול מודול/מנשך בסיס הקיים כיום במערכת אחזר מטמכים, שיקרא מודול/מנשך BASE ושמכיל רשימה מדורגת. מטרת הניסוי לבדוק אם הפתרון המוצע שונה מהפתרון המקורי כיום וכיצד, ואת בהיבטים של זמני תגובה ומספר המטמכים שיש לסרוק, וכן לבחון מספר ממדדים סובייקטיביים.

מדובר במאגר שבו המטמכים תיוגו מראש בצורה ידנית במסגרת הניסוי נמדדו הן זמני התגובה וכוכנות התשובות (נתונים אובייקטיביים), והן, לדוגמה, הנוחות והיעילות עבור משתמשים (נתונים סובייקטיביים). בניסוי נמצא כי המודל אכן יכול לשרת משתמשים בצורה יעילה יותר ממודלים אחרים שקיים בשטח.

## מערך הניסוי

מערך הניסוי כלל ביצוע שימושים חיפוש שונים על ידי קבועה של שיטות משתחפות. השימוש הופעלן באמצעות מנשך BASE ובמנשך TTL&NISI. לצורך הניסוי הוקם מאגר מידע (צר-הכל כארבע מאות מטמכים) شامل מסמכי תשובה למשימות החיפוש שהוגדרו. מאגר המטמכים המקורי מכיל מיליון מטמכים שונים ידנית, ומתחום נגזרו המטמכים שהוועדים לשיטות תשוכות לשילופות שהוגנו מראש. הניסוי השווה שתי צורות של הצגת התשובות בעזרת שאלות מוכנות מראש, ואוסף תשוכות שמתאים להן הקטגוריות במנשך TTL&NISI הוצגו על ידי מבנה הטקסטונומיה הארונית.

שימוש החיפוש הוגדרו מראש בהתאם על גישות של דומאיס ועמיתו (Dumais, Cutrell and Chen, 2001)

## מודל TTL&NISI (Taxonomy Tree & LCC&K Snippet)

moatzu CAN מודל הצגת תשובות חדש, אשר יסיע למשתמש לאחרזר את מכוקשו בזמן קצר יותר, ולמקד אותו במסמכים הרלוונטיים עכורים. האתגר הוא למצוא דרך מיטבית לצגת המידע הכללי קטגורית, על מנת לסייע למשתמש בתהיליך ניפוי מספר גדול של תשובות (Mordechai, Drori & Frank, 2007).

הרעין המרכזי מבוסס על חשיפת מידע-על הקים במאגר במסגרת הממשק עצמו, ועל שימוש בו לצורך ארגון וסידור רישימת התוצאות על פי עקרונות HFC. מידע-על זהה מהוות למעשה "מבחן לוגי" על רישימת התוצאות, וכך לסייע במשימות חיפוש מחקריות שבהן משתמשים בקטגוריות כדי להחליט על המשך: להתגלל בתוצאות, לעזן שליפה ועוד. כמו כן מוצע לשפר את יכולת תקציר המסמרק על מנת לסייע למשתמש להחליט אם המסמרק רלוונטי, ללא הצורך לפתח אותו לבדיקה מעמיקה יותר.

המודל החדש מציע שיפור השיטה לצגת תוצאות חיפוש מסוימים כך שתכלול שילוב של:

1. קיבוץ המטמכים על פי קטגוריות - הקטגוריות מבוססות על תיוגים שנעשו ידנית על סמך ערכי טקסטונומיה ארגונית, הקטגוריות מסודרות בכמה פנים, כאשר כל פן הוא ענף ראשי בעץ הטקסטונומיה. ענפי העץ יזרגו על פי אלגוריתם שיפורוט בהמשך.
2. הצגת השורות הרלוונטיות לערכי החיפוש וכן קטגוריות ומילוט מפתח בכל תקציר מסמרק (Snippet) על פי הגדרת מנשך LCC&K. אבטיפוס המישם את מודל TTL&NISI מוצג באור 1.

- יתרונות המודל המוצע:
- המשתמש מקבל מידע לגבי פיזור המטמכים בין הקטגוריות והפנים.
- המבנה מסיע להמקד בקטgorיה מסוימת ללא הצורך בסדרת כוורות המטמכים.
- תקצירים המטמכים עוזרים להבין במקרה הצורר את תכולת המסמרק, כדי להחליט אם המסמרק רלוונטי.

## דירוג הקטגוריות

כדי לקבוע את סדר הקטגוריות בעץ הטקסטונומיה נשתמש בשיטה מקובלת לדירוג מטמכים:  $tfidf$  (term frequency inverse document frequency).

שיטה זו מבוססת על מודל וקטורי עבור כל מסמרק והוא שיטת דירוג פופולרית, מכיוון שהיא נוחה לפשטה ול מהירה (Baeza-Yates & Ribeiro-Neto, 1999).

moatzu CAN לכצע **הקובלה** של המונחים בנוסחה המקורית (מלים במטמכים) למונחים שלנו (נושאים בשליפה), כדלהלן:

### שכיחות המילה במטמרק ~ שכיחות הנושא בשליפה

ולכן: חסכנות קטגוריה  $Ci$  בשליפה:

# לאמרים מחקריים

ביצוע הניסוי. הזמן שנחון לנבדקים לבצע המשימות היה בלתי מוגבל. נוכנות התשובה או התשובות אומתיה בזמן הניסוי על ידי עורך הניסוי.

## משימות הניסוי

מערך המשימות הכיל ארבע משימות חיפוש מתוך עולם התוכן של המשתמש. שתיים מהמשימות בעלות אופי מוקד-איתור מסמך אחד שבו נמצא המידע הדරושים, ושתי משימות חיפוש הן בעלות אופי מחקרי - איתור מידע הנמצא ביותר מסמך אחד. מדובר במשרד ממשלתי, ולכן ניתן לחפש במפורש את התרחישים שהוצעו ואת תוכן ערכי השlipה במסימות מכל מקום, נדגים את הסגנון ואופי השאלה בעזרת משימות חיפוש בתחום התרבות והפנאי:

- דוגמה לשאלת חיפוש מוקדית: באיזו שנה זכה יזהר כהן באירוויזיון?
- דוגמה לשאלת חיפוש מחקרית: מצא דוגמאות למסמכים שבהם מוזכר הקשר בין השחקן מומי מושונוב לתיאטרון הימה, ובאיזה הצגות הוא שיחק?

מיוקם המסמכים הרלוונטיים בכל אחת מארבע המשימות מפורט בטבלה 1. סדר השlipה אוזן בצורה אקראית, כך שככל משימה בוצעה על ידי משתמשים שונים בדרכים שונות.

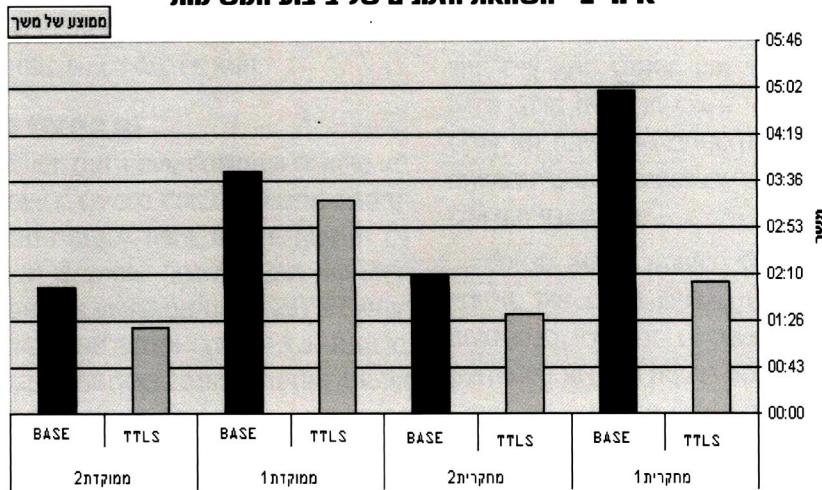
במחקר שבוצע בצורה דומה, ההגיון בגישה זו אומר לנטרל את השונות הקיימת ביכולות החיפוש של משתמשים ברמות שונות, ולאפשר בכך לפחות בצורה נקייה ולא רעשים את שלב סקירת התשובות של המשתמשים, הכספי לשיפור. מוגבלה זו תוטר, כמובן, ביחס מלא של המודל. סוג המשימות נלקחו מתוך עכודה אמיתית ויצור הנעה יכולת להשתמש במידע המוצעו שלו ביצוע המשימה.

יעילות החיפוש וuschiwot הרצן נבדקו תוך שימוש בשני טדים: נתונים אובייקטיביים: זמן וnocnoot התשובה; נתונים סובייקטיביים: נוחות,יעילות עבור המשתמש, הרגשת ביטחון, שביעות רצון ומידת הרלוונטיות של ריכבי המידע. לאחר השלמת המשימות בכל אחד מהמנשכים, מילא המשתמש שאלון הנתונים מולאו בצורה מוחשכת ואוחסנו במסד נתונים. כמו כן אספה המערכת נתונים ליום (Log) באופן המאפשר לבצע ניתוח מדויק של זמני התגובה ושל תיעוד הפעולות שכיצעו הנבדקים במהלך הניסוי.

## סמל הניסוי

לצורך הניסוי הוקם מאגר מידע שככל מסכים ומסמכים רלוונטיים. הניסוי נערכ בצורה בדידה בין עורך הניסוי לנבדק-אחד על אחד". ניתנה אפשרות לבצע משימה דמה לפני

איור 2 - השוואת הזמינים של ביצוע המשימות



## ריכוז התוצאות

**MEAN** - הערך הממוצע שהתקבל.

**SD (Standard Deviation)** - סטיית התקן של הממוצע ביחס לכל התוצאות. ככל שסטיית התקן קטנה יותר, פיזור הנתונים קטן יותר וההתוצאה הממוצעת משמעותית יותר.

**P (Probability)** - הסתברות, שבסקרה זה היא אחוז הסיכוי למקרים שהחוקר מוכן להסתכן בו ולקבל את התוצאות.

סבלה 1  
מיוקם המסמך/כימ הרלוונטיים מתוך הרשימה המקורית

משימה	מיוקם המסמך/כימ הרלוונטיים מתוך הרשימה המקורית
ספקטט 1	13 מתוך 33
ספקטט 2	6 מתוך 95
ספקטט 3	99,100,103 מתוך 109
מחקרית 2	63,73,85,90,91,92 מתוך 119

## הסבר לתוצאות

כשמשתמש سورק את תחילת הרשימה ואינו מוצא בה את מבקשו, הוא פונה לעזרת עץ הטקסטונומיה. כשהמשתמש מופיע בקטגוריה המבוקשת או לכזו שנראית לו רלוונטי, הוא נעזר בתקציר המסמכים המאפשר לו להחליט, בצורה מהירה יחסית, אם המספר רלוונטי לשאלת החיפוש, הן על פי ה קטגוריות המשייבות לו והן על סמך שורות מהטקסט שבחן מופיעות מילול החיפוש. כתוצאה לכך, מספר המסמכים שהמשתמש נכנס אליהם עד שהוא מוצא את מבקשו, קטן יותר ותהליךஇitor יותר התשובה מתקצר בצורה משמעותית.

## סיכום ומסקנות

כדי להתגבר על הקושי להתמודד באופן יעיל עם מספר גודל של תשובות, הוצע מודל TTLS המשלב מספר טכניקות להציג תשובות, וככלון עץ טקסטונומיה ותקציר מסמר משופר מטרת האבטיפוס שנבנה על סמך המודל היה להראות כי המודל שהוצע אכן פועל את הבעיה העיקרית. באבטיפוס בוצע ייסוי הן של זמני הביצוע וכוכנות התשובות, והן של בחינת הנחות והיעילות עבור המשתמשים. בניסוי נמצא כי המודל אכן יכול לשרת משתמשים בצורה יעילה יותר ממודלים אחרים שקיים בשיטה.

כדי להניבר את יעלותו של המודל ולענות על צרכים נוספים של המשתמשים, מומלץ להשקיע בשיפורם, הרחבות ובבדיקות שונות במודל ובאבטיפוס. מחקר עתידי יבחן את הממשק הנוכחי מול משתמשים נוספים, כך שהיקף התוצאות המתקבלות יהיה גדול יותר, ונתן יראה להפיק בכלים סטטיסטיים מסקנות ורחבות יותר. כיוון אחר למחקר הוא יצירה אוטומטית של הקטגוריות שנacen להיום הקשורות בצורה ינית למסמכים. בנוסף, ניתן יהיה לבחון את התאמיות המתקבלת בין הקטלוג הידני לבין הקטלוג האוטומטי שיעשה שימוש באלגוריתמים קיימים לקטלוג מסמכים בעולמות ידע נתוניים.

לסייע, השיטה המוצעת לדירוג ענפי עץ הטקסטונומיה היא ייחודית. זאת בנוסף לחלקים שונים ממנשך המשתמש שכר הצענו בעבר, כל אחד שלעצמם, כגון הצגת עץ הטקסטונומיה, הצגת פנים או הצגת תקציר מסמר על פי מודל LCC&K. השימוש המוצע של המרכיבים לעיל במודל TTLS ותהליך ההערכה עם משתמשים אמיתיים, מופיע את התחרמתה המוצגת במאמר זה בתחום לאחר המידע.

מעין בטבלה 2 בה מוצגות תוצאות מבחן T (המשמש לבדיקת השערת על הבדלים של שתי קבוצות שאינן חיליות זו בזו), ניתן לראות כי בכלל אחד מהפרמטרים, מלבד כמות המידע שהוצגה, יש יתרון ברור למנשך TTLS.

**טבלה 2**  
תוצאות מבחן T לבדיקת שונות בין המנשכים

משתנה נבדק	טובות P<0.05		מנשך MEAN (SD)
	מבחן T	מנשך TTLS BASE	
הרגשות נוחות בתחילת	יש	4.41 (0.76)	2.71 (1.05)
האיתור	אין	3.45 (0.62)	3.40 (1.06)
רלוונטיות המידע	יש	4.33 (0.54)	2.50 (0.89)
הנלווה לຄורתת	יש	2.06 (0.60)	2.63 (0.80)
הטעה במידע הנלווה	לא		
לຄורתת	יש	4.46 (0.53)	2.78 (0.86)
השימוש ביחסון בזמן	לא		

ההשוואת תוצאות הזמן (איור 2), עולה בכירור כי ממוצע זמני הביצוע של המשימות בעזרת מנשך TTLS נמוך יותר מממוצע זמני הביצוע של המשימות בעזרת מנשך BASE. BASE כמשמעותו במשימות חיפוש מחקריות, הופיע בזמן גודל אף יותר. במשימות המוקדמות הופיע נגבי בין 13-44 אחוז, ואילו במשימות המתקדמות הופיע נגבי בין 39-240 אחוז.

## מספר המסמכים שנפתחו

תקציר המספר המשופר נועד לסיעו למשתמש להחליט אם המספר רלוונטי ללא הצורך להיכנס אליו. כל כניסה לממסך שהמשתמש ביצע ונרשמה לקובץ היום, ובוצעה השוואת בין שתי השיטות. ניתן לראות שבשימוש במושך BASE היה צורך לפחות מסמכים מאשר בעבודה עם מושך TTLS. לעומת זאת, המידע המבוקש במושך BASE נפתחו רק 6.7 מסמכים ממוצע, לעומת זאת, במסמכים ואילו ב-TTLS נפתחו רק 2.2 מסמכים ממוצע. מניתוח תוצאות השאלונים עולה, באופן כללי, שהרובה המוחלט של המשתמשים (סך-הכל 77 אחוז) יעדיפו להשתמש במושך TTLS לעיתים קרובות ואףלו תמיד.

## מקורות

- Baeza-Yates R. & Ribeiro-Neto B. (1999). *Modern Information Retrieval*. Addison-Wesley, New York.
- Dumais S. & Chen H.(2001). Hierarchical classification of web content, *SIGIR 2000*. Athens, Greece, ACM Press, New York, pp. 256-263.
- Dumais S. T., Cutrell E. & Chen H.(2001). Optimizing search by showing results in context, *CHI 2001*. Seattle, WA, USA, pp. 277-284.
- Drori O. & Alon N.(2003). Using documents classification for displaying search results list, *Journal of Information Science*, 29(2), 97-106.

- Drori O. & Tamir E.(2005). *Display of Search results in Hebrew: a comparison study between Google and LCC&K interface*, *Journal of Information Science*, 31(3), 164-177.
- Drori O.(2000). *Refinement of Search in Information Systems and Improving Presentation of Results for End Users, as Efficient Tools for Decision Making Process*, Ph.D. Thesis, The Hebrew University, Jerusalem.
- Hearst M. A.(2006). Clustering versus faceted categories for information exploration, *Communications of the ACM*, 49(4), 59-61.
- Ivory M.Y., Yu S. & Gronemyer K.(2004). search result exploration: a preliminary study of blind and sighted users' decision making and performance, *CHI 2004*. Vienna, Austria, ACM Press, pp. 1453-1456.
- Kaki M.(2005). Findex: search result categories help users when document ranking fails, *CHI 2005*. Portland, Oregon, ACM Press, pp.131-140.
- Mordechai V., Drori O., & Frank A. J.(2007). *Grouping Search Results by Organizational Taxonomy Using LCC&K Interface*, Technical Report No. 2007-5, Leibniz Center for Research in Computer Science, School of Computer Science and Engineering, Hebrew University of Jerusalem.