

## ניסוחי שאילתות במערכות לאיחזור מידע לצרכים אקדמיים / שלי חן-ארידור

### תקציר

חיפוש מידע הוא תהליך מורכב למרות היותו חלק משגרת היומיום. הצלחת חיפוש מידע הוגדרה במחקרים שונים באופנים שונים ואין הגדרה פורמלית להצלחה של חיפוש מידע. המחקרים מראים שחיפוש מוצלח תלוי במספר מרכיבים, ביניהם: מבנה המידע (כגון: טקסט מלא, thin metadata, thick metadata), מערכת חיפוש המידע ומחפש המידע (המשתמש).

עבודות מחקר קודמות הצביעו על הצלחה מוגבלת של חיפוש מידע (על-פי הגדרות שונות) בתחומים שונים הן בווב, והן במאגרי מידע ביבליוגרפיים ובקטלוגים של ספריות. בהתאמה, מחקרים רבים בתחום חיפוש המידע מציעים שיטות וטכניקות לשיפור מערכות חיפוש המידע הן באופן ביצוע החיפוש על ידי המערכת (למשל, אלגוריתמים או אפשרויות חיפוש מתקדמות וייחודיות) והן ברמת עיצוב ממשק המשתמש (למשל גישה ברורה ונוחה לאינדקסים). יש להניח כי מערכות חיפוש המידע ימשיכו להשתכלל ולאפשר חיפושים יעילים בעתיד ואף חיפושים המתבססים על משמעות מונחי החיפוש ועל הקשרים אסוציאטיביים. אך ככל הנראה, בטווח הנראה לעין, המשתמש ימשיך להיות גורם מרכזי בתהליך חיפוש המידע באמצעות מערכות אלו. רק הוא יהיה מסוגל להגדיר את צורכי המידע שלו ולהעריך את התוצאות (רלוונטיות, סמכות, כיסוי וכו').

שימוש מושכל במערכות המידע (כגון: ניסוח נכון של צורך המידע, בהתאמה למערכת איחזור המידע, שימוש באופציות חיפוש ייחודיות ואו מתקדמות, שימוש בטקטיקות חיפוש יעילות) יכול להביא לשיפור מימדי הצלחה של תהליך החיפוש (כגון: רלוונטיות התוצאות, הזמן המוקדש לחיפוש). לכן ישנה חשיבות רבה למחקרים הבודקים את מרכיב המשתמש בתהליך חיפוש המידע בכלל ובהצלחת החיפושים מהיבט המשתמש בפרט. עבודת מחקר זו שייכת לקבוצת המחקרים בתחום התנהגות חיפוש המידע, העוסקת בניתוח שאילתות החיפוש ורלוונטיות התוצאות, מנקודת מבטו של המשתמש.

המחקר התמקד בחיפוש מידע לצורך כתיבת עבודה אקדמית בקרב חוקרים צעירים. בהתאם, אוכלוסיית המחקר כללה שישה-עשר תלמידי מחקר לתארים מתקדמים מהמחלקה ללימודי-מידע של אוניברסיטת בר-אילן, אשר ערכו

חיפוש-מידע לצורך כתיבת עבודה אקדמית. הציפייה היתה כי סטודנטים אלו, מתוקף השכלתם ואולי אף מהרקע המקצועי שלהם, יהיו בעלי יכולות חיפוש מידע גבוהות מהממוצע באוכלוסיה ויכלו לכן לתת תובנות ממוקדות יותר לקשיים בתהליך החיפוש. כתוצאה יזוהו מקומות אפשריים לשיפור התהליך. נקודות קושי שיתגלו באוכלוסיה זו בודאי יופיעו גם באוכלוסיות בעלות אוריינות חיפוש נמוכה יותר.

המחקר השתמש במתודולוגיה איכותנית כדי לקשר ולבדוק את ניסוחי השאלות לסטודנט ספציפי, כנגד נושא/שאלת המחקר של עבודה אקדמית ורלוונטיות התוצאות. נכון להיום, ישנם מספר מועט של מחקרים איכותניים שכאלו. בנוסף, מחקר זה כלל מרכיב ייחודי של השוואת התקדמות החיפוש והצלחה בין מאגרי מידע שונים עבור אותו משתמש. כל משתמש חיפש מידע (מוגבל מראש ל- session חיפוש יחיד) עבור אותו נושא בשלושה מאגרים. המאגרים עבורם נערך המחקר היו: קטלוג הספרייה של אוניברסיטת בר-אילן, גוגל-סקולר ומאגרי פרוקווסט. השוואה זו איפשרה להעריך את השפעת המנשק ואפשרויות החיפוש הייחודיות של כל מאגר על הצלחת החיפוש.

בהתבסס על המידע שנאסף מניסוחי השאלות (בעזרת תוכנת הקלטות מסך וקול), מהשיח של המרואיינים ומנתוני השאלונים ובעזרת מתודולוגיה איכותנית, מחקר זה ביקש לענות על השאלות הבאות:

- האם יש מאפיינים ייחודיים לשאלות שנכשלו?
- מהי האבולוציה ששאלות עוברות במהלך החיפוש?
- האם יש הבדלים בהתייחסות המשתמשים למאגרים השונים?

בהתייחס לשאלת המחקר הראשונה, נמצאו מספר גורמים לשאלות שנכשלו (בהגדרה, אלו שלא הניבו אף פריט מידע שהמשתמש בחר לשמור או הביע עניין בו בקול רם):

- שליטה לא מספקת באנגלית.
- קשיים בניווט עד למציאת הטקסט המלא של פריטי המידע המבוקשים.
- חוסר מיומנות בשימוש בפונקציות חיפוש יחודיות של מנשקים.

- קושי בתרגום נושא המחקר ושאלת המחקר למונחי חיפוש מתאימים וקושי במציאת מונחים נרדפים, צרים, רחבים, המבטאים את ההיבטים הנושאים השונים של שאלת המחקר.
  - היעדר אסטרטגיית חיפוש שיטתית (למשל חזרה על שאילתות, אי-שימוש במטאדטה). ביצוע חיפושים ללא תכנון מראש, מבלי ללמוד מתוצאות החיפושים (השאילתות) הקודמים.
- לשאלה האם ישנה תבנית ברורה של התפתחות שאילתות במהלך חיפוש, ממצאי המחקר לא הניבו תבניות ברורות. נהפוך הוא, בנוסף לבעיות טכניות (כגון: שפה, הבנת המנשקים) אובחן בבירור היעדר אסטרטגיית חיפוש שיטתית, היעדר שימוש במשוב משאילתות קודמות המבוססת על תוצאות. בנוסף, התופעה אינה ייחודית למאגר מידע ספציפי.
- בהשוואת התייחסות המשתמשים למאגרים השונים, נמצאו ההבדלים הבאים:
- שביעות הרצון מתוצאות החיפוש במאגרים: שביעות הרצון בקטלוג היתה נמוכה לעומת שאר המאגרים.
  - זמן: השקעת הזמן בקטלוג היא קטנה בהשוואה לזמן שהושקע בחיפושים בגוגל-סקולר ובמאגרי פרוקווסט. קשה לקבוע אם השקעת הזמן הקצרה היא אחת הסיבות לחוסר שביעות הרצון מהתוצאות או ההיפך.
  - שפת החיפוש: הסטודנטים נטו לערוך חיפושים בעברית (שפת אם) במאגרים בעלי מנשק בעברית. למרות נוחות החיפוש בשפת-האם, במרבית המקרים, חיפושים אלו היו פחות מוצלחים. על מתכנני מנשקי מערכות איחזור המידע לקחת בחשבון נטייה זו, כאשר מתרגמים מנשק של מערכת איחזור מידע לשפת האם כאשר אין פריטי מידע בשפה זו.
  - שאילתות ואסטרטגיות חיפוש:
    - לביצוע שאילתות חיפוש בקטלוג יש מאפיינים המבחינים מאגר זה מהשניים האחרים: 1. שימוש במטאדטה בשאילתות הראשונות. 2. חיפוש היבט נושאי אחד בשאילתה הראשונה. 3. אחוזי הצלחת שאילתות הנמוכים פי שלושה מגוגל-סקולר.
  - אוצר המילים המשמשים ב-session החיפוש:
    - אוצר המילים חולק לשלוש קבוצות: מונחי ליבה, בהם השתמשו בכל המאגרים; מונחים המשותפים לשני מאגרים; מונחים בהם השתמשו רק במאגר אחד, נקראו מונחים ייחודיים.

- שימוש במונחים ייחודיים בכל מאגר, אשר בחלקם הניבו תוצאות רלוונטיות למשתמש, מרמזות על שימוש לא שיטתי במונחי החיפוש.
- בגוגל-סקולר ובפרוקווסט משתמשים ביותר מונחי-חיפוש משותפים מאשר עם הקטלוג.
- למונחי החיפוש יש יעילות משתנה באיחזור פריטי מידע רלוונטיים, באותו מאגר (בטקטיקות חיפוש שונות) ובמאגרים שונים. כלומר, יש חשיבות מרכזית בשימוש נכון במנשק.

ממצאי המחקר לעיל מעידים על קיומם של מספר מכשולים בתהליך חיפוש. בהמשך, המחקר מעלה רעיונות ודרכי פעולה ליעול תהליך החפוש. על המשתמשים להבין את מבנה המידע וכיצד מערכות איחזור מידע פועלות, למשל, הבחנה בין חיפוש בטקסט מלא וחיפוש במטאדטה. יש להקנות למשתמשים ידע ספציפי בטכניקות לאחזור מידע (למשל: שימוש בפקודות קינון, שימוש במטאדטה) ומיומנות בשימוש בטקטיקות חיפוש מגוונות (למשל מיקוד החיפוש, הכללת החיפוש, שימוש במונחים צרים/רחבים/נרדפים), ובכלל זה היכרות עם אופרטורים שונים (בעיקר האופרטור OR) ואופני פעולתם. בנוסף, ללמדם שימוש שיטתי במונחי החיפוש ושילובם באסטרטגיות חיפוש מתוכננות בתהליך אינטראקטיבי של שיפור החיפוש, בהתבסס על התוצאות מהשאלות הקודמות.

חשוב להעלות למודעות המשתמשים כי למערכות איחזור מידע שונות אופני פעולה שונים וכי אסטרטגיית החיפוש צריכה להיות מותאמת לאפשרויות ויכולות המערכת. יש חשיבות לידיעת ההגדרות הספציפיות בכל מערכת והיכרות עם אינדקסים ייחודיים או אפשרויות הגבלה ייחודיות. כמו-כן חשוב להתייחס לאופן אחסון המידע וסוג המידע עליו עורכים את החיפוש (טקסט מלא, thin metadata, thick metadata). שימוש מושכל במנשק איחזור המידע יעלה את אחוזי הצלחת השאלות.

יש להקנות מיומנויות גם בהקשר של הגעה למקורות המידע עצמם, בעיקר הכוונה ליכולת להגיע לטקסט המלא (למשל על ידי לימוד ההבדל בין מקורות חופשיים ברשת לבין מקורות בתשלום ומקורות בתשלום המסופקים חינם לסטודנטים על ידי הספרייה). הבנת מסלולי הגישה למקורות המידע יקצרו את הזמן המוקדש לאיתורם ויאפשר הגעה למספר רב יותר של פריטי מידע רלוונטיים.

על המשתמשים בעלי שפת-אם עברית להעשיר את אוצר המילים בתחום צורך המידע שלהם (מונחים מקצועיים, צרים, רחבים, נרדפים) לטובת שאילתות יעילות. יש ללמדם להשתמש בכלי עזר שפתיים ספציפיים (כגון מילונים ותיזאורוסים). כמו-כן, יש להדגיש את החשיבות בשימוש במונחים רלוונטיים מתוך תוצאות החיפוש. אין לסמוך רק על שיפור מערכות החיפוש. הממצאים במחקרנו, מצביעים על הצורך בחינוך לאוריינות מידע כדי לאפשר חיפוש מידע יעיל באופן עצמאי על ידי המשתמש וללא תיווך של מידענים. אוריינות מידע כוללת מיומנויות רבות, ביניהן: היכולת לחפש ולאתר את המידע הדרוש באופן יעיל, יכולת להעריך את המידע ואת מקורות המידע באופן ביקורתי, היכולת להשתמש במידע ביעילות להשגת מטרה ספציפית.

הקניית מיומנויות אלו תיתן מענה לסוגיות שעלו במחקר זה ותהווה יישום של המסקנות. מומחי מידע (כגון מידענים, ספרנים, אשר מומחיותם מתבססת על רקע תיאורטי, ידע ייחודי ונסיון מקצועי) יכולים להעניק את הכלים הדרושים למשתמשים לביצוע חיפושים יעילים ולפתח אוריינות מידע. המלצה זו נתמכת בממצאים אמפיריים המעידים כי אוריינות מידע היא גורם משמעותי בהצלחה בלימודים. יתר על כן, אנו מאמינים שאפשר וצריך לשפר באופן מהותי את יעילות החיפושים. הדבר נכון לא רק לגבי סטודנטים ללימודי מידע, אלא גם לגבי אוכלוסיית החוקרים הכללית באקדמיה ואף לגבי כלל האוכלוסייה. כולם זקוקים לכישורי חיפוש אפקטיביים לצורך מילוי צורכי המידע השונים והמגוונים בחיי היומיום, בעבודה, לצרכי פנאי וכדמי'.

**מספר מערכת:**

**1206012**

**מספר מיון בספרייה:**

**025.4 חן.נסו תשע"א**