

## סוג מנוע החיפוש והרגלי אחזור המידע האישי: "חלונות" לעומת לינוקס

עופר ברגמן      רות בייט-מרום      רפי נחמias      נעה גרדוביץ      סטיב וויטקר  
אוניברסיטת שפילד      האוניברסיטה הפתוחה      אוניברסיטת תל אביב      אוניברסיטת שפילד      אוניברסיטת שפילד  
b.ofer@shef.ac.uk      ruthbm@open.ac.uk      nachmias/noahgrad@post.tau.ac.il      s.whittaker@shef.ac.uk

### Type of Search Engine and Personal Information Retrieval: "Windows" versus Linux

Ofer Bergman      Ruth Beyth-Marom      Rafi Nachmias      Noa Gradovitch      Steve Whittaker  
Sheffield University      The Open University      Tel Aviv University      Tel Aviv University      Sheffield University  
of Israel

#### Abstract

Two of the main ways by which personal files can be retrieved are navigation and search. The *search everything* approach asserts that improved search engines will lead to preference for search over navigation in PIM. The aim of this study was to test this assumption (as part of a bigger research). Participants were 47 computer programmers who completed a questionnaire concerning their retrieval habit with both Windows XP OS (older generation search engine) and Linux OS (advanced search engine). There was little evidence supporting the search everything claim. First, regardless of the search engine there was a strong navigation preference: on the average, users estimated that they used navigation for 61% with Windows and 67% with Linux. Second, although participants estimated that they used search significantly more with Linux ( $M=20\%$ ) than with Windows ( $M=6\%$ ), the increase in search was not a result of a decrease of navigation, but of the fact that there is no "recent documents" option in Linux. Third, most participants indicated that when they did not remember their file location, they used search as a last resort. Finally, there was no evidence that using an advanced search engine led participants to rely less on their hierarchal file organization.

**Keywords:** Personal information management, search, navigation.

#### תקציר

שתי הדרכים העיקריות בהן משתמשים על מנת לאחזר קבצים אישיים הן ניווט וחיפוש. גישת החיפוש טוענת ששיפור מנועי החיפוש יביאו להעדפת חיפוש על ניווט בניהול מידע אישי. המחקר נועד לבדוק טענה זו. השתתפו במחקר 47 מתכנתים שהשתמשו בשתי מערכות הפעלה חלונות XP אשר מנוע החיפוש שלו נחשב למיושן ולינוקס בעל מנוע חיפוש מתקדם. המשתתפים מילאו שאלון בו דיווחו על הרגלי האחזור שלהם בשתי מערכות ההפעלה. נמצאו מעט עדויות אשר תומכות בטענת גישת החיפוש: (א) הערכת הנבדקים לגבי שתי מערכות ההפעלה מעידה על העדפה ברורה של ניווט על פני חיפוש: בממוצע 61% מהאחזורים בחלונות ו-67% מהאחזורים בלינוקס התבצעו באמצעות ניווט. (ב) אומנם המשתתפים העריכו שהם מחפשים בלינוקס (20% בממוצע) יותר מבחלונות (6% בממוצע), אולם עלייה זאת בחיפוש לא באה על חשבון הניווט אלא משום שבלינוקס לא קיימת אפשרות האחזור "קבצים אחרונים". (ג) מרבית המשתמשים העידו ש כאשר הם מתקשים לזכור את מיקום הקבצים, הם מבצעים חיפוש רק כמוצא אחרון. (ד) לא נמצאו תימוכין לכך ששימוש

במנועי חיפוש מתקדמים מוביל לירידה בהסמכות על שמירה היררכית. המאמר מסביר את הממצאים בהבדלים בין ניהול מידע אישי וניהול מידע באינטרנט, ומפנה לגישה הסובייקטיבית-למשתמש אשר מציעה מספר ממשקים אפשריים לשיפור הניווט.

**מילות מפתח:** ניהול מידע אישי, חיפוש, ניווט.

## מבוא

ניהול מידע אישי היא פעילות שבה אדם מארגן ושומר את פריטי המידע שלו כדי לאחזר אותם בעצמו בזמן מאוחר יותר. בהקשר של המחשב האישי פריטים אלו יכולים להיות קבצים, הודעות דוא"ל ומועדפים. ניהול מידע זה נעשה בעזרת מערכות לניהול מידע אישי, כגון מערכת ההפעלה, תוכנת הדוא"ל ודפדפן האינטרנט. שתי הדרכים העיקריות אשר ניתן לאחזר באמצעותן פריטי מידע הן ניווט וחיפוש. ניווט הוא תהליך בו משתמשים "מטיילים" במורד היררכיית התיקיות עד שהם מגיעים למיקום בו נמצא הפריט המבוקש, ואז מוצאים אותו בתוך מיקום זה. חיפוש הוא תהליך בו המשתמשים יוצרים שאילתא המכילה חלק מהמאפיינים של הפריט המבוקש, ואז בוחרים את הפריט הרלוונטי מתוך רשימת התוצאות. ניווט, בניגוד לחיפוש, דורש אחסון היררכי. כלומר, משתמשים צריכים ליצור תיקיות או מחיצות (directories) ולאחסן את פריטי המידע "בתוכן" כהכנה לאחזור עתידי. המחיצות הוצגו לראשונה בשנות השישים והתיקיות בשנות השמונים ובמשך העשורים שעברו מאז נעשה באחסון היררכי שימוש כמעט בלעדי ללא שינויים משמעותיים.

הביקורת נגד השיטה ההיררכית מתייחסת לקשיים הן בשלב האחסון והן בשלב האחזור. משתמשים עלולים להתקשות בבחירת התיקייה בה הם מאכסנים את פריט המידע משום שניתן לסווג אותו לקטגוריות שונות (Dumais et al., 2003; Malone, 1983). בזמן האחזור הם עלולים להתקשות בזכירת הקטגוריה שבה בחרו לאחסן את הפריט (Kidd, 1994; Lansdale, 1988; Whittaker & Sidner, 1996). בניסיון לפתור קשיים אלו מוצגת גישה החיפוש שנועדה להחליף את שיטת הניווט ולהקל (לדעת חסידיה) הן בשלב האחסון והן בשלב האחזור. גישה זו אינה מצריכה קבלת החלטות באשר לקטגוריית אחסון שכן היא פוטרת את המשתמש מהעול של שמירה בתיקיות ומאפשרת למשתמש להיעזר בכל מאפיין שהוא זוכר לגבי פריט המידע בכדי לאחסן אותו (Cutrell, Dumais, & Teevan, 2006; Dourish et al., 2000).

הביטחון בגישה החיפוש כפתרון לקשיים של גישה הניווט באחסון מידע ובאחזור, שאב את כוחו כנראה מהדומיננטיות של מנועי חיפוש באינטרנט (Cutrell, Robbins, Dumais, & Sarin, 2006). כדי לעודד את השימוש בחיפוש גם במערכות של ניהול מידע אישי, נעשו ניסיונות חוזרים ונשנים לשפר את מנוע החיפוש, חלקם ויתרו על היררכיות תיקיות לחלוטין והסתמכו על כלי חיפוש באפן בלעדי (Fertig, Freeman, & Gelernter, 1996b) *Lifestreams*, *Canon Cat* (Raskin, 2000) ו-*Presto* (Dourish, Edwards, LaMarca, & Salisbury, 1999). למרות זאת, בתחום ניהול המידע אישי הראיות מעידות דווקא שמשתמשי מחשב מעדיפים "באפן גורף" לנווט לקבצים שלהם על פני ביצוע פעולת חיפוש (Barreau & Nardi, 1995). בתגובה ענו מצדדי גישה החיפוש שמנועי החיפוש שהיו קיימים בתקופה בה נעשה המחקר (ובעיקר מנוע החיפוש של מערכת ההפעלה DOS) הם "איטיים, קשים להפעלה, פועלים רק על שמות (ולא תוכן)" וכן ששיטת האינדוקס שלהם אינה טובה (Fertig, Freeman, & Gelernter, 1996a. p. 67). לדעתם של פרטיג ועמיתיו שיפור מנועי החיפוש תביא להעדפה של חיפוש על ניווט גם בניהול מידע אישי.

בעשר השנים שחלפו מאז מאמרם של פרטיג ועמיתיו חלה התקדמות טכנולוגית אדירה במנועי החיפוש. השיפור התבטא בעיקר בתחומים עליהם הם הצביעו כלקויים: מהירות החיפוש, חיפוש טקסט מלא (כלומר מילים בתוך קובץ ולא רק בשם שלו), ואינדקסים משוכללים. אחד ממנועי החיפוש המתקדמים ביותר בקריטריונים אלו הוא זה של מערכת ההפעלה לינוקס. זה הזמן לבחון את טענתם של פרטיג ועמיתיו ולבדוק האם מנוע חיפוש משופר יעלה באופן משמעותי את החיפוש על חשבון הניווט באחזור מידע. המחקר הנוכחי נועד לבחון סוגייה זו.<sup>1</sup>

<sup>1</sup> כחלק מסדרה של מחקרים עם מטרת מחקר זהה.

## שאלות המחקר

1. כיצד מעדיפים משתמשי מחשב למוצא קובץ, על ידי ניווט או חיפוש?
2. האם הזמינות של מנוע חיפוש מתקדם מעלה את השימוש בחיפוש על חשבון ניווט?
3. מתי המשתמשים מבצעים חיפוש?
4. האם השימוש במנועי חיפוש מתקדמים מוביל לירידה בהסתמכות על שמירה היררכית?

## מתודולוגיה

במחקר השתתפו 47 מתכנתים אשר משתמשים במערכת ההפעלה לינוקס בעבודתם ובחלונות XP בביתם. מערכת ההפעלה של לינוקס משתמשת כאמור במנוע חיפוש מתקדם לעומת זה של מערכת ההפעלה חלונות שנחשב למיושן במיוחד (Farina, 2005; Lowe, 2006; Pogue, 2004). במחקר השתמשנו במתודולוגיה תוך אישית within subject design, כלומר השוונו את העדפות כל משתתף בשני תנאים עם לינוקס (מנוע חיפוש מתקדם) ועם חלונות XP (מנוע חיפוש מיושן). המחקר עשה שימוש בשאלון אשר הכיל שלושה חלקים: שאלות לגבי אחזור בחלונות, שאלות לגבי אחזור בלינוקס (מחצית המשתתפים קיבלו שאלון המתחיל בחלונות ומחציתם בלינוקס), ושאלות רקע.<sup>2</sup> אחד המשתתפים המרכזיים במחקר הוא הערכת המשתמש את אחוזי האחזורים באמצעות חיפוש מתוך סה"כ אחזורי הקבצים (להלן אחוזי החיפוש המוערך). הערכה זו עברה תיקוף באמצעות השוואה לאחוזי אחזורי החיפוש בפועל. לצורך זאת נלקח מדגם של 190 פקודות אחזור במערכת ההפעלה לינוקס מ-11 נבדקים. נמצא שאחוזי החיפוש בפועל ( $M = 15\%$ ,  $SD = 13\%$ ) היה נמוך במעט מאחוזי החיפוש המוערך ( $M = 20\%$ ,  $SD = 14\%$ ).<sup>3</sup>

## תוצאות

כדי לענות על שתי שאלות המחקר הראשונות התבקשו הנבדקים להעריך אחוזי השימוש שהם עושים בכל אחת מאפשרויות האחזור שמערכת ההפעלה מציעה להם בטבלה שהסתכמה ב 100%, וזאת באופן נפרד לגבי שתי מערכות ההפעלה. תוצאות הערכות אלו מופיעות בטבלה 1:

טבלה 1. הערכת אחוז השימוש בכל אחת מאפשרויות האחזור – ממוצע (סטיית תקן)

סה"כ	אחר	Linux GUI	מסמכים אחרונים	קיצורי דרך	חיפוש	ניווט	
100%	1% (7%)	-	14% (18%)	18% (19%)	6% (5%)	61% (24%)	חלונות
100%	1% (2%)	2% (5%)	-	10% (13%)	20% (14%)	67% (18%)	לינוקס

### כיצד מעדיפים משתמשי מחשב לניווט אל הקבצים שלהם, על ידי ניווט או חיפוש?

הערכת הנבדקים לגבי שתי מערכות ההפעלה מעידה על העדפה ברורה של ניווט על פני חיפוש. טבלה 1 מראה שבמערכת ההפעלה חלונות העריכו הנבדקים בממוצע שהם משתמשים בניווט ב-61% מסך כל האחזורים שהם מבצעים ואילו באמצעות חיפוש הם מבצעים רק כ-6% מאחזורים שהם מבצעים ( $t(46)=10.87$ ,  $p<0.01$ ). במערכת ההפעלה לינוקס אחוזי הניווטים המוערך היה 67% בממוצע, לעומת אחוזי חיפוש מוערך של 20% בממוצע ( $t(46)=14.72$ ,  $p<0.01$ ).

### האם השימוש במנוע חיפוש מתקדם מביא ליותר שימוש בחיפוש?

המחקר הראה שאחוזי החיפוש המוערך במערכת ההפעלה לינוקס ( $M = 20\%$ ) היה גבוה באופן מובהק מזה של מערכת ההפעלה חלונות ( $M = 6\%$ ) ( $t(46)=6.1$ ,  $p<0.01$ ). אולם, טבלה 1 מראה שהעלייה באחוזי החיפוש לא באה כתוצאה מירידה באחוזי הניווטים אלא משום שבלינוקס לא

<sup>2</sup> לקבלת השאלון המלא אנא פנו לנעה גרדוביץ.

<sup>3</sup> לא התאפשר לנו לבצע מתאמים משום שהשאלונים היו אנונימיים, אולם ניסוי אחר מסדרה זו הראה מתאם גבוה ביותר של 0.94 בין הערכת האחוזי החיפוש ובין החיפוש בפועל.

קיימת אפשרות של "קבצים אחרונים". כאשר חישבנו את ההעדפה היחסית בכל מערכת הפעלה על ידי חיסור אחוז החיפוש המוערך מאחזור הניווטים המוערך לא נמצא הבדל בין מערכות הפעלה  $t(34)=1.47, p>0.05$ .

### מתי המשתמשים מבצעים חיפוש?

כדי לענות על שאלה זו התבקשו המשתתפים להשלים את המשפט "אני מאחזר קבצים בעזרת חיפוש כאשר \_\_\_\_\_". 89% מהמשתתפים השיבו שהם השתמשו בחיפוש כאשר הם לא זכרו את מיקומו של הקובץ במערכת הפעלה חלונות ו-77% השיבו כך במערכת לינוקס. במבחן t למתאמים תלויים לא נמצא הבדל משמעותי בין שתי מערכות הפעלה  $t(46)=0.23, p>0.05$ . ממצאים אלו מעידים שהשימוש באפשרות החיפוש נעשה בעיקר במצב של חוסר ברירה – כאשר אין למשתתפים אפשרות לנווט אל הקובץ המבוקש.

### האם השימוש במנועי חיפוש מתקדמים מוביל לירידה בהסתמכות על שמירה היררכית?

מצדדי גישת החיפוש טוענים שמנועי חיפוש טובים יותר יביאו לירידה בהסתמכות על השמירה ההיררכית. לטענתם משתמשים אשר ברשותם מנועי חיפוש טובים יתרשלו בשמירה על הקבצים כיוון שהם יוכלו לחפש אותם בקלות. התרשלות זאת תתבטא בהכרח בעלייה באחוז האחזורים בהם המשתמש מתקשה לזכור מיקומו של הקובץ מתוך סהכ' האחזורים שהוא מבצע. אולם כאשר התבקשו משתתפי המחקר להעריך אחוז זה התשובה הממוצעת שלהם הייתה זהה לגבי האחזורים בחלונות ( $M = 26\%, SD = 26\%$ ) ובלינוקס ( $M = 26\%, SD = 20\%$ ). ממצאים אלו אינם תומכים בטענת גישת החיפוש ששימוש במנועי חיפוש מתקדמים מוביל לירידה בהסתמכות על שמירה היררכית.

### דיון

ממצאי המחקר אינם תומכים בטענות גישת החיפוש: המשתתפים העריכו שאת מרבית האחזורים שלהם הם מבצעים באמצעות ניווט ורק מיעוט האחזורים באמצעות חיפוש. ממצאים אלו עקביים עם ממצאים קודמים (Boardman & Sasse, 2004; Capra & Pérez-Quñones, 2005; Teevan, 2004; Alvarado, Ackerman, & Karger, 2004). השימוש במנוע החיפוש המתקדם לינוקס לא העלה את אחוז האחזורים היחסי מעבר לשימוש במנוע החיפוש המיושן של חלונות XP; משתתפי המחקר העידו שהם מבצעים חיפוש בעיקר במצב של חוסר ברירה – כאשר הם אינם זוכרים את היכן הקבצים שלהם נמצאים; והשימוש במנוע חיפוש טוב יותר לא הביא לעלייה בהערכה של אחוז האחזורים בהם המשתתפים מתקשים לזכור את מיקומו של הקובץ.

חיפוש הוא השיטה העיקרית בה מאחזרים המשתמשים מידע ברשת, אולם בניהול מידע אישי נראה שחיפוש משמש רק כמוצא אחרון. נראה שישנו הבדל מהותי בין שתי סביבות המידע, אשר גורם, להבדלים בהתנהגות האחזור. בניהול מידע אישי אותו המשתמש מאחסן ומאחזר את המידע. המשתמש מארגן את המידע בצורה סובייקטיבית שמתאימה לצרכיו ומייצגת את הבנתו לגבי המידע (Jones, Phuwanartnurak, Gill, & Bruce, 2005). המשתמש מכיר את מבנה התיקיות שלו והכרות זו עוזרת לו למצוא את הקובץ הדרוש. ברגמן (2006) הציע מספר הסברים תיאורטיים מתחום הפסיכולוגיה הקוגניטיבית להעדפת ניווט על חיפוש.

בשנים האחרונות הושקעו משאבי ענק בשיפור מנועים לחיפוש מידע אישי. מאמצים אלו לא היו לשווא שכן משתמש אשר אינו מוצא את הקבצים שלו זקוק לכל העזרה שהוא יכול לקבל וטוב שתהא יעילה. אולם נראה שהחלום שחיפוש יחליף את הניווט לאחזור מידע אישי הוא חלום שווא. לפיכך אנו מציעים להפנות חלק ממאמצים אלו לטובת שיפור הניווט. הגישה הסובייקטיבית-למשתמש אשר פותחה במיוחד עבור מערכות לניהול מידע אישי מציעה מספר ממשקים אפשריים לשיפור הניווט (Bergman, Beyth-Marom, & Nachmias, 2003, in print). השפעתם תיבדק במחקר עתידי.

### Acknowledgment

This research was partly fund by the European Union Marie Curie Grant, TOK 30008.

## מקורות

- ברגמן ע' (2006). השימוש במאפיינים סובייקטיביים במערכות לניהול מידע אישי. עבודת דוקטורט, אוניברסיטת תל אביב, תל אביב, ישראל. עמ' 80-86.
- Barreau, D. K., & Nardi, B. A. (1995). Finding and reminding: file organization from the desktop. *SIGCHI Bulletin*, 27(3), 39-43.
- Bergman, O., Beyth-Marom, R., & Nachmias, R. (2003). The user-subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology*, 54(9), 872-878.
- Bergman, O., Beyth-Marom, R., & Nachmias, R. (in print). The user-subjective approach to personal information management systems design –evidence and implementations. *Journal of the American Society for Information Science and Technology*.
- Boardman, R., & Sasse, M. A. (2004). "Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 583-590). Vienna, Austria: ACM Press.
- Capra, R. G., & Pérez-Quñones, M. A. (2005). Using Web Search Engines to Find and Refind Information *Computer*, 38(10), 36-42.
- Cutrell, E., Dumais, S. T., & Teevan, J. (2006). Searching to eliminate personal information management. *Communications of the ACM*, 49 (1), 58-64.
- Cutrell, E., Robbins, D. C., Dumais, S. T., & Sarin, R. (2006). Fast, Flexible Filtering with Phlat: Personal Search and Organization Made Easy. In *CHI 2006 Conference on Human Factors in Computing Systems* (pp. 261-270). Montreal, Canada: ACM Press.
- Dourish, P., Edwards, W. K., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., et al. (2000). Extending document management systems with user-specific active properties *ACM Trans. Inf. Syst.* , 18(2), 140-170.
- Dourish, P., Edwards, W. K., LaMarca, A., & Salisbury, M. (1999). Presto: an experimental architecture for fluid interactive document spaces *ACM Transactions on Computer-Human Interactions*, 6(2), 133-161.
- Dumais, S. T., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R., & Robbins, D. C. (2003). Stuff I've seen: a system for personal information retrieval and re-use. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 72-79). Toronto, Canada: ACM Press.
- Farina, P. A. (2005). *A Comparison of Two Desktop Search Engines: Google Desktop Search (Beta) vs. Windows XP Search Companion*. Paper presented at the 21st Computer Science Seminar.
- Fertig, S., Freeman, E., & Gelernter, D. (1996a). "Finding and reminding" reconsidered *SIGCHI Bulletin*, 28(1), 66-69.
- Fertig, S., Freeman, E., & Gelernter, D. (1996b). Lifestreams: an alternative to the desktop metaphor. In *Conference companion on Human Factors in Computing Systems: Common Ground* (pp. 410-411). Vancouver, British Columbia, Canada: ACM Press.
- Jones, W., Phuwanartnurak, A. J., Gill, R., & Bruce, H. (2005). Don't take my folders away! Organizing personal information to get things done. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. Portland, Oregon: ACM Press.
- Kidd, A. (1994 ). The marks are on the knowledge worker In *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence* (pp. 186-191). Boston, Massachusetts, United States ACM Press.
- Lansdale, M. W. (1988). The psychology of personal information management. *Applied Ergonomics*, 19(1), 55-66.
- Lowe, M. (2006). *Evaluation of desktop search applications*. Sydney, Australia: Kalio.

- Malone, T. W. (1983). How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems, 1*, 99-112.
- Pogue, D. (2004, 21 Oct. 2004). Google Takes On Your Desktop. *The New York Times On The Web*.
- Raskin, J. (2000). *The humane interface: new directions for designing interactive systems*. Boston: ACM Press/Addison-Wesley Publishing Co.
- Teevan, J., Alvarado, C., Ackerman, M. S., & Karger, D. R. (2004). The perfect search engine is not enough: a study of orienteering behavior in directed search. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 415-422). Vienna, Austria: ACM Press.
- Whittaker, S., & Sidner, C. (1996). Email overload: exploring personal information management of email. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems: Common Ground* (pp. 276-283). Vancouver, British Columbia, Canada: ACM Press.