

# פלאים של מנועי החיפוש - המקרה של חיפוש מידע על חוקר האינטרנט אנדרי ברודר

יהודית בר אילן

אנדרי ברודר (Andrei Broder) הוא חוקר אינטרנט ידוע. מאחר שלא היה לו דף בית (homepage) באינטרנט, קשה היה למצוא מידע עדכני עליו, במיוחד כיוון שהחליף כמה מקומות עבודה בעשר השנים האחרונות. במאמר זה נבחן את המידע המתקבל עליו באמצעות מנועי חיפוש. מאחר שב-Web לא מייחסים חשיבות למועד עדכון הדפים, תוצאות החיפוש מכילות הן מידע מיושן והן מידע עדכני וקשה מאוד להחליט איזה מידע עדכני. בנוסף לכך, נתאר את מאמצי המתחרים בתחרות חיפוש מידע באינטרנט במאי 2005 למצוא מידע על תפקידו ומקום עבודתו של אנדרי ברודר. בגלל הקושי במציאת המידע, הם נקטו בשיטות שונות המתאימות למודל ה-berrypicking של Bates (1989).

## מבוא

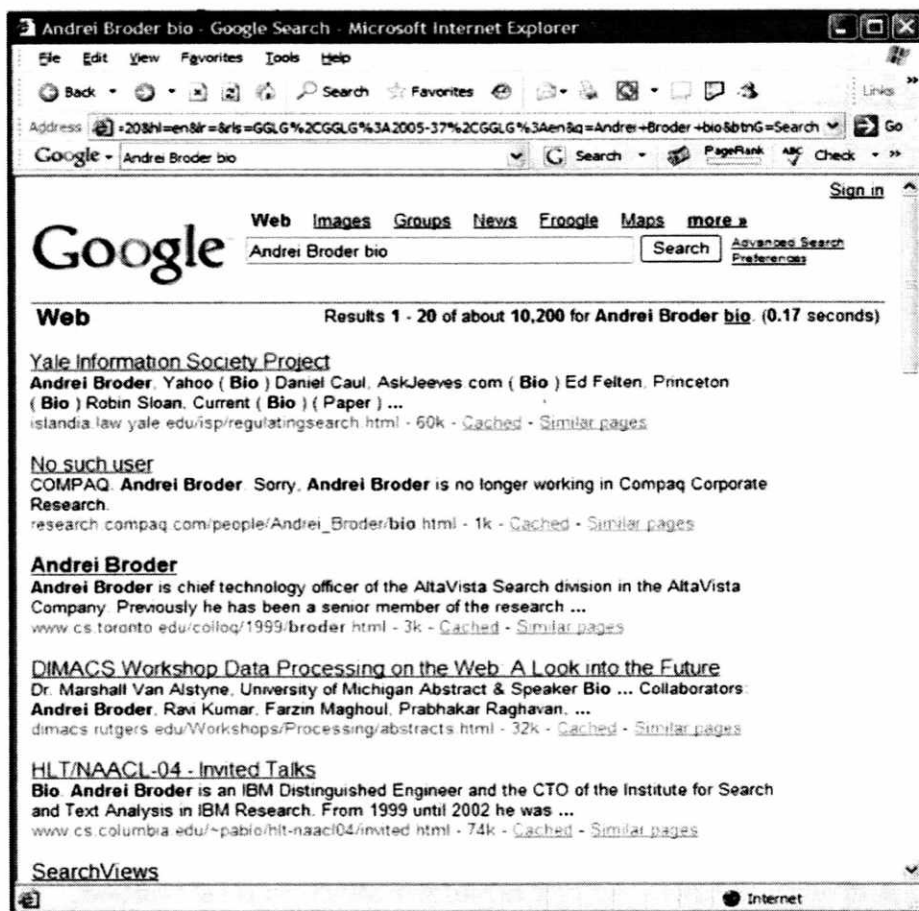
ה-Web (מארג בעברית תקנית, אך נמשך להשתמש במונח Web) הוא מקור מידע מרכזי עבור אוכלוסיית המדינות המפותחות. הוא מכיל מיליארדים רבים של דפים, כך שאחת הבעיות המרכזיות איתה צריך להתמודד היא עומס המידע. כדי להתגבר לפחות באופן חלקי על בעיה זו, אנחנו נעזרים במנועי חיפוש. מנועי החיפוש מדרגים את התוצאות בעזרת אלגוריתמים ייחודיים של דירוג התוצאות, כדי שהמשתמשים יראו בראש רשימת התוצאות את התשובות ה"רלוונטיות" ביותר לשאלותיהם. גורמים רבים משפיעים על הדירוג, אך בניגוד למאגרים ביבליוגרפיים, לא ניתן לדרג את דפי ה-Web לפי מועד פרסומם, כיוון שלא ניתן לקבוע בצורה מהימנה את מועד פרסומם של דפי ה-Web.

כאשר אנחנו מחפשים מידע עדכני על אנשים, יש חשיבות רבה למועד פרסום המידע, במיוחד כאשר מחפשים מידע על עבודתו הנוכחית של אדם שהחליף כמה מקומות עבודה בתקופה האחרונה. ברודר (Broder, 2002) סיווג את שאילתות המשתמשים לשלושה סוגים: שאילתות מידע, שאילתות ניווט ושאילתות טרנזאקציה. חיפוש מידע על מקום עבודתו של אדם ותפקידו מתבצע על ידי שאילתת ניווט, כאשר אנחנו מצפים למצוא את דף הבית של האישיות. כפי שמסביר ברודר (Broder, 2002), אם מישהו מעוניין במידע על חוקר מדעי המחשב Donald Knuth, הוא יריץ את השאילתה Donald Knuth וירצה להגיע לדף הבית של החוקר (<http://www-cs-faculty.stanford.edu/~knuth>). אך מה קורה כאשר לאישיות המבוקשת אין דף בית (כמו לברודר עצמו - מקרה מובהק של הסנדלר שהולך יחף!). במקרה כזה עלינו להסתמך על מידע שמתפרסם על ידי מקורות אחרים. אך מקורות אלה עלולים להיות בלתי מדויקים או בלתי מעודכנים. במאמר זה נתחקה אחרי "עקבות ה-Web" של אנרי ברודר. ד"ר ברודר עבד בחברת DEC SRC בין השנים 1984 ו-1988. מידע קודם אודותיו מתייחס לתקופה שלפני ה-Web, אך ניתן למצוא פרטים גם על תקופה זו, למשל ב-Yahoo! Media Relations, 2005; Yale Information Society Project, 2005. במחצית השנייה של 1988 חברת DEC התמזגה עם חברת Compaq ובין השנים 1998 ו-1999 היה ברודר חוקר בכיר ב-Compaq SRC (ב-2002 Compaq התמזגה עם HP ושמו של Compaq SRC שונה ל-HP SRC, אך בתקופה זו אנדרי ברודר כבר לא עבד ב-Compaq/HP). באוגוסט 1999 מונה ל-Chief Technology Officer ב-AltaVista. הוא נשאר ב-AltaVista (מנוע חיפוש) עד פברואר 2002 - אז עבר ל-IBM ומונה ל-Distinguished Engineer and the CTO of the Institute for Search and Text Analysis. מאז נובמבר 2005 הוא משמש Research Fellow and Vice President of Emerging Search Technology at Yahoo! Research (ראה

ד"ר יהודית בר אילן - המחלקה ללימודי מידע, אוניברסיטת בר-אילן. [barilaj@mail.biu.ac.il](mailto:barilaj@mail.biu.ac.il)  
 תרגום של המאמר: False web memories: A case study on finding information about Andrei Broder  
 אשר פורסם בכתב העת המקוון: First Monday, [http://firstmonday.org/issues/issue11\\_9/barilan/index.html](http://firstmonday.org/issues/issue11_9/barilan/index.html)

.(Yahoo! Media Relations, 2005) אנדרי ברודר הועסק במקומות עבודה רבים בשמונה השנים האחרונות. נדגים זאת בעזרת תוצאת השאילתה *Andrei Broder bio*, שהונחה ב-1 במאי 2006 למנוע החיפוש Google (ראו תרשים מספר 1). ארבע התוצאות הראשונות מרמזות על מקומות עבודה שונים: IBM, Compaq, Yahoo. התוצאה השנייה נראית מעניינת במיוחד: "No such user". בהמשך נבחן את מיקומו של דף זה במנועי חיפוש שונים במשך שלוש שנים. בנוסף לכך, ננתח את מאמצי איתור המידע אודות תפקידו וכתובת מקום עבודתו העדכני של אנדרי ברודר על ידי 49 המשתתפים בתחרות חיפוש המידע שנערכה במאי 2005 (<http://www.afeka.ac.il/siteFiles/1/68/571.asp>).

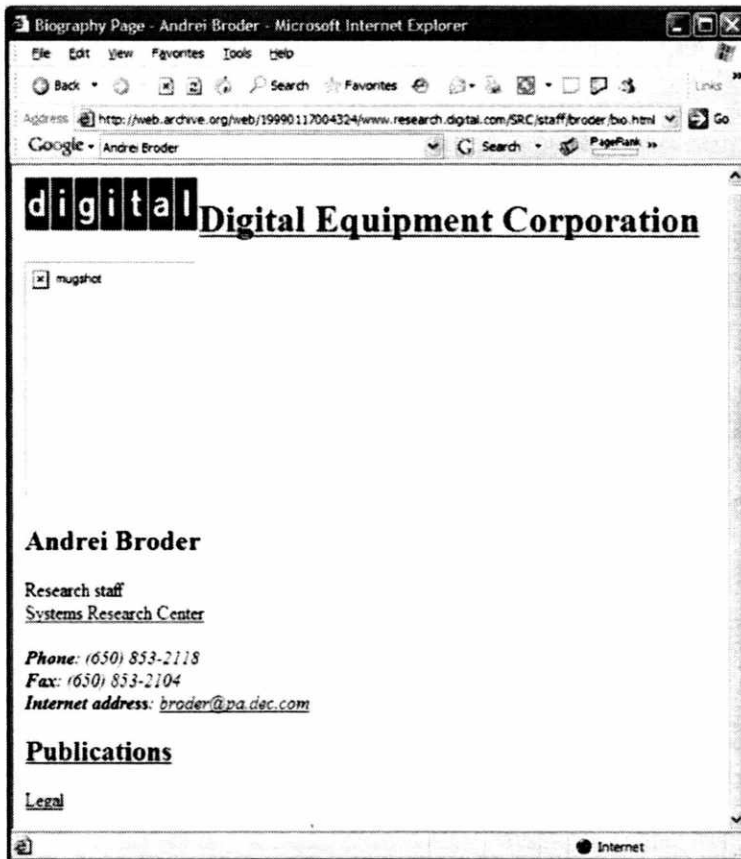
**תרשים 1:** תוצאות ראשונות של השאילתה "Andrei Broder bio" שהתקבלו מ Google ב-1 במאי 2006



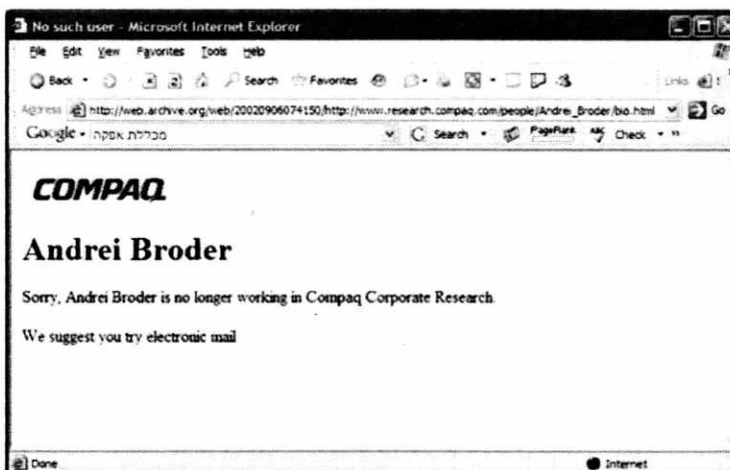
## שינויים במיקומו של הדף "No such user" בתוצאות החיפוש

אנדרי ברודר עזב את חברת Compaq באוגוסט 1999. דף זה נמצא בשתי כתובות אינטרנט חליפיות:  
 (1) [http://research.compaq.com/people/Andrei\\_Broder/bio.html](http://research.compaq.com/people/Andrei_Broder/bio.html)  
 (2) [http://www.research.digital.com/people/Andrei\\_Broder/bio.html](http://www.research.digital.com/people/Andrei_Broder/bio.html)  
 בארכיון האינטרנט (<http://archive.org>) ניתן למצוא עותק של דף מינואר 1999, כאשר אנדרי ברודר היה עדיין מועסק בחברת Compaq (ראו תרשים 2). בארכיון האינטרנט אין עותקים משנת 2000, יש שלושה עותקים רשומים מ-2001, אך במקומם מופיע העותק מספטמבר 2002 (ראו תרשים 3). הדף לא השתנה מאז, אבל השרת אינו נגיש לפחות מאז מרץ 2006, למרות שמנועי החיפוש: Google, Yahoo!, Ask ו-Exalead עדיין מיפתח דף זה ביוני 2006 (בנובמבר 2006, Google, Live Search (MSN) ו-Exalead ממפתחים את הדף). עותק המטמון של Google הוא מדצמבר 2005.

תרשים 2: <http://www.research.digital.com/SRC/staff/broder/bio.html> - 17 בינואר, 1999



תרשים 3: הדף "No such user" כפי שנשמר ע"י ארכיון האינטרנט ב-6 בספטמבר, 2002



תרשים 4: תוצאות החיפוש של Google ב-22 בינואר, 2003



בתרשים 4 אפשר לראות את תוצאות החיפוש עבור השאלתה Andrei Broder בתחילת 2003. חזרנו שוב על החיפוש במאי 2003 והפעם הזנו את השאלתה במנועי החיפוש: Teoma, Google, AlltheWeb, Alta Vista. במועד זה הדף "No such user" היה תוצאה שלישית ברשימת התוצאות של Google, ראשונה ב-Alta Vista, מספר 11 ב-Teoma ולא הופיע כלל בין שלושים התוצאות הראשונות עבור AlltheWeb. חזרנו על החיפוש שוב במאי 2005. במועד זה היה הדף הנדון התוצאה השנייה ב-Google, תוצאה שלישית ב-Teoma, לא היה בין מאה התוצאות הראשונות של Yahoo! וגם לא בין חמישים התוצאות הראשונות של MSN (במשך השנים השתנו גם מנועי החיפוש המובילים). במאי 2006 הדף הנדון הופיע כתוצאה שמינית ברשימת התוצאות של Google, אך לא היה בין 1,000 התוצאות של Yahoo! - מצגי עבור חיפוש כלשהו (גם כאשר יש יותר מ-1,000 תוצאות, ו-Google ו-Yahoo! מוסר מידע על 1,000 URL-ים בלבד). למרות שדף זה לא הופיע בחיפוש הספציפי, היה עדיין מופתח ב-Yahoo! - הדף הופיע כתוצאה מספר 24 גם ברשימת התוצאות של Ask (ה"יורש" של Teoma), ולא מופתח במועד זה על ידי MSN.

מעניין לדעת מדוע דף זה, שאינו קיים ככל הנראה ואינו מכיל שום מידע רלוונטי ממופתח על ידי מנועי החיפוש ומופיע במקום כה גבוה בתוצאות החיפוש 6.5 שנים אחרי שאנדרי ברודר עזב את חברת Compaq. הדף מכיל 13 מילים בלבד, כל אחת מהמילים Andrei ו-Broder מופיעה פעמיים בטקסט ופעם נוספת ב-URL. בדקנו גם את מספר הקישורים המובילים לדף זה (גם מספר הקישורים אל הדף משפיע על הדירוג). Google לא מצא אף קישור, אך ידוע ש-Google אינו מדייק בחיפושים מסוג זה (ראו Forums.SearchengineWatch, Bar-Ilan, 2002 או 2004). Yahoo! החזיר שמונה דפים עם קישור לדף הנדון. כך שנראה שמספר הקישורים המובילים אל הדף אינה הסיבה למיקומו הגבוה. אנחנו משערים שהגורמים העיקריים המשפיעים הם הופעת מונחי החיפוש ב-URL והופעתם התכופה יחסית בטקסט עצמו.

## מציאת מידע עדכני על אנדרי ברודר

נניח שאנו במאי 2005 והוטלה עלינו משימה למצוא "מה תפקידו הנוכחי וכתובת מקום עבודתו של חוקר האינטרנט Andrei Broder (הניסוח המדויק של השאלה כפי שהופיעה בתחרות חיפושי מידע באינטרנט - [http://www.afeka.ac.il/search-form-1.asp?info\\_id=607](http://www.afeka.ac.il/search-form-1.asp?info_id=607)). כפי שצינו קודם, זאת שאלתת ניווט טיפוסית (Broder).

כך שכדי לענות עליה, היה המשתמש הממוצע פונה ל-Google (מנוע החיפוש הפופולארי ביותר, ראו comScore, 2005) בשאלתה *Andrei Broder* והיה מעיין בשלוש התוצאות הראשונות (לפי בדיקות שערכה חברת Enquiro ב-2005 משתמשים מתייחסים לתוצאות החיפוש הראשונות בלבד). במקרה הנוכחי, אסטרטגיה זו אינה משיגה את המבוקש (ראו תרשים 5). התוצאה הראשונה לא מספקת כל מידע על מקום עבודתו של אנדרי ברודר, התוצאות השנייה והשלישית נראות סותרות - מתוך הכותרת של התוצאה השנייה ניתן להסיק שאנדרי ברודר אינו עובד ב-Compaq בעוד שהתוצאה השלישית רומזת לכך שהוא עובד של חברת Compaq. אם המשתמש בוחר לעיין בתוצאה השלישית ואינו מסתפק בתקציר, ימצא: "Our collaborators include... Andrei Broder (AltaVista)...". כלומר, ניתן להסיק שאנדרי ברודר עובד בחברת AltaVista. אבל למעשה, במאי 2005 ברודר היה מועסק יותר משלוש שנים בחברת IBM. ייתכן, שהמשתמש הממוצע היה פשוט מוותר בשלב זה, אך המשתתפים בחצי הגמר של תחרות החיפוש היו עקשנים יותר, בפרט משום ששאלה זו נראתה פשוטה יחסית לשאלות האחרות ([http://www.afeka.ac.il/search-form-1.asp?info\\_id=607](http://www.afeka.ac.il/search-form-1.asp?info_id=607)).

**תרשים 5:** תוצאות החיפוש של Google עבור השאלתה *Andrei Broder* ב-1 במאי, 2005



חצי הגמר של אליפות החיפוש השנייה נערכה ב-1 במאי 2005 במעבדות המחשבים של מכללת אפקה. לשלב זה הוזמנו שישים מתחרים שהיו אמורים לענות נכונה על מספר רב ככל האפשר של שאלות מתוך 12 שאלות בתוך שתיים. שישים המתחרים עברו את המוקדמות שנערכו באמצעות האינטרנט. התשובות בכל השלבים נבדקו על ידי שופטי התחרות. יותר מחמישים אחוז מהמשתתפים בחצי הגמר היו סטודנטים ו/או עובדי ה"י-טק. יותר משני שלישים מהם היו בעלי תואר אוניברסיטאי או למדו לקראת תואר; רוב המתחרים דיווחו שהם מבליים לפחות עשרים שעות בשבוע באינטרנט. ארבעת המתחרים הטובים ביותר הגיעו לשלב הגמר שנערך במסגרת כנס INFO2005. בכל המחשבים במעבדה הותקנה בידיעת המתחרים תוכנת מעקב, שתיעדה את מהלכי המתחרים במשך התחרות. (ראו בתקנון התחרות: [http://www.afeka.ac.il/info\\_clean.asp?info\\_id=614](http://www.afeka.ac.il/info_clean.asp?info_id=614). רשימת השאלות מופיעה ב: [http://www.afeka.ac.il/search-form-1.asp?info\\_id=607](http://www.afeka.ac.il/search-form-1.asp?info_id=607)).

כאמור, שאלה מספר 5 היתה: "מה תפקידו הנוכחי וכתובת מקום עבודתו של חוקר האינטרנט Andrei Broder". 49 מתוך שישים המשתתפים ניסו להתמודד עם שאלה זו. תשובותיהם מסוכמות בטבלה 1. כמעט רבע מהמשיבים הגיעו למסקנה המוטעית שאנדרי ברודר מועסק על ידי חברת החיפוש Alta Vista, למרות שבאותו זמן הוא כבר עבד יותר משלוש שנים בחברת IBM.

טבלה 1: תשובות המתחרים (N=49)

תשובה	מספר משתתפים	% משתתפים
תשובה נכונה – YAHOO!	16	32.7%
Alta Vista	12	24.5%
IBM Watson	8	16.3%
תפקיד נכון, חסרה כתובת	6	12.2%
IBM research	4	8.2%
כתובת נכונה, חסר תפקיד	2	4.1%
IBM + Princeton	1	2.0%

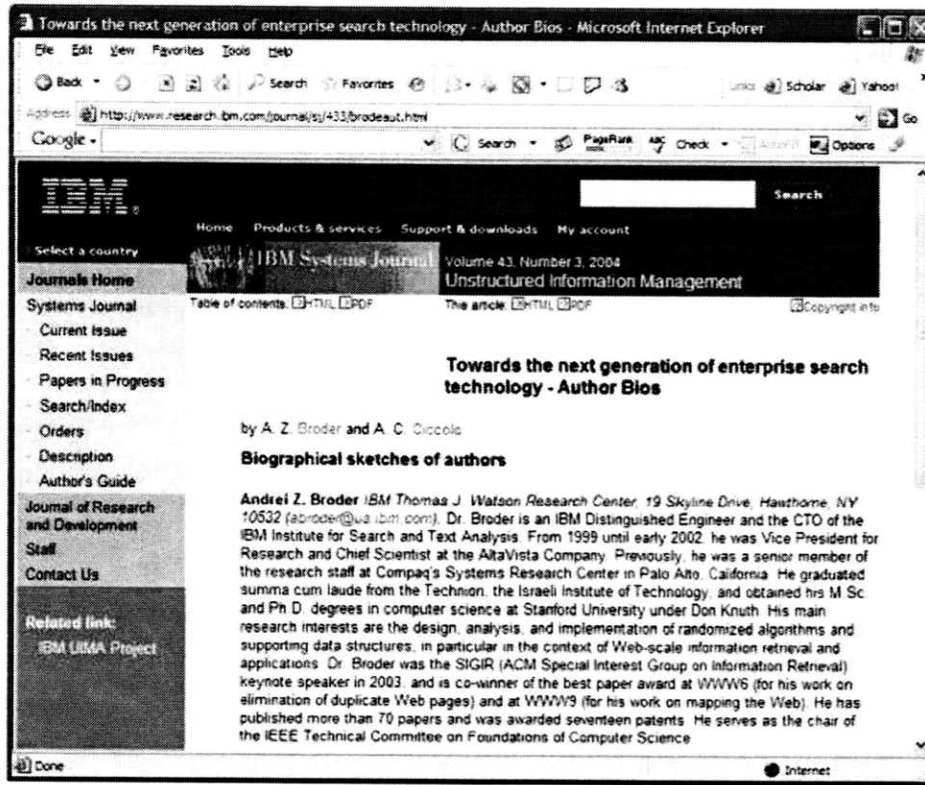
כצפוי, השאלתה הראשונה של רוב המתחרים היתה פשוט Andrei Broder. למרבית הצער, לא פעלה תוכנת המעקב באופן שווה בכל המחשבים, לכן יש בידינו דיווח מפורט רק על 33 מתוך 49 מתחרים אשר התמודדו עם שאלה זו. בשארית המאמר נתבסס רק על מהלכי החיפוש שביצעו 33 במשתתפים הנ"ל. טבלה 2 מציגה את התפלגות שאלתת הפתיחה.

טבלה 2: התפלגות שאלתת הפתיחה (N=33)

שאלתת הפתיחה	מספר משתתפים	% משתתפים
Andrei Broder	20	60.6%
Andrei Broder ומילים נוספות	7	21.2%
"Andrei Broder" ומילים נוספות	3	9.1%
"Andrei Broder" (כביטוי)	2	6.1%
אנדרי ברודר	1	3.0%

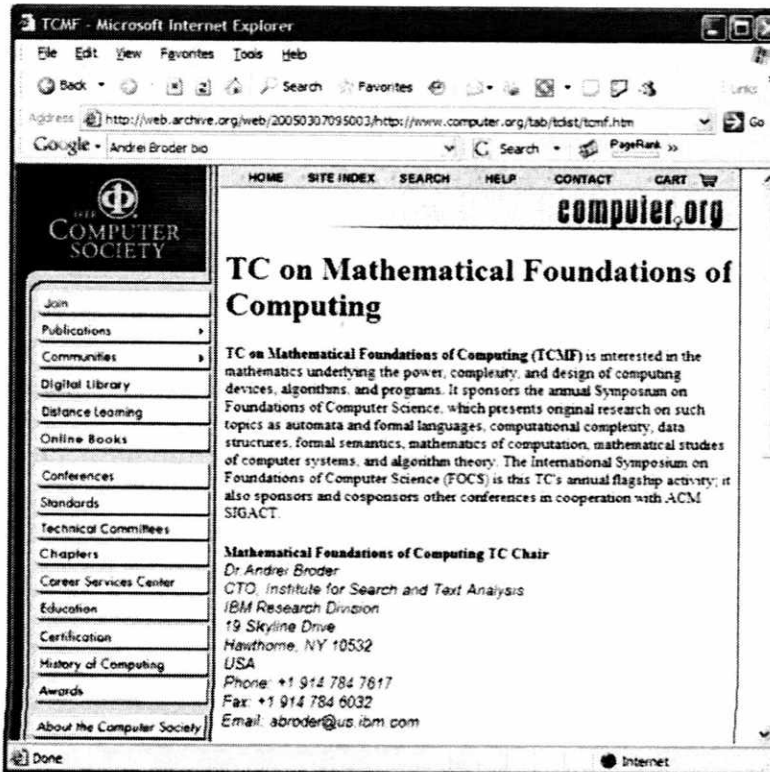
חמישה משתתפים בשלב הראשון של החיפוש בחרו לחפש את המחרוזת המדויקת "Andrei Broder". בדרך כלל, המטרה של חיפוש ביטוי (המונחים בתוך הגרשיים) היא להביא לתוצאות ממוקדות יותר. במקרה זה אסטרטגיה זו לא הובילה לתוצאות המקוות, מכיוון ששמו המלא של Andrei Broder הוא Andrei Z. Broder, ובשני דפי ה-Web שבו היה ניתן למצוא תשובה מלאה לשאלה, שמו הופיע עם ראשי התיבות של שמו האמצעי (ראו תרשימים 6 ו-7). מתוך המשתתפים אשר החלו את החיפוש עם שאלתת שהכילה את שמו כביטוי (סך הכל חמישה משתתפים) הצליח רק משתתף אחד לענות באופן מלא על השאלה. מצד שני, אף אחד מהמשתתפים לא הגיע למסקנה המוטעית שהוא עובד בחברת Alta Vista. טבלה 3 מציגה את ה-URLים עליהם ביססו המשתתפים את תשובותיהם.

תרשים 6: <http://www.research.ibm.com/journal/sj/433/brodeaut.html>



תרשים 7: <http://www.computer.org/tab/tclist/tcmf.htm>

דף זה אינו נגיש יותר, הוחזר בארכיון האינטרנט (<http://archive.org>), העותק מ-7 במרץ, 2005



## טבלה 3: הכתובות השכיחות ביותר עליהן התבססו המשתתפים כדי לענות על השאלה

נבחר מס' פעמים	המידע ב-URL	כתובת ה-URL	1
14 (29%)	תשובה מלאה	<a href="http://www.research.ibm.com/journal/sj/433/brodeaut.html">http://www.research.ibm.com/journal/sj/433/brodeaut.html</a>	1
4 (8%)	תפקיד	<a href="http://www.mitacs.ca/interchange/bios/border.html">http://www.mitacs.ca/interchange/bios/border.html</a>	2
3 (6%)	researcher at IBM Watson	<a href="http://www.mail-archive.com/nlpatumd@yahoogroups.com/msg00008.html">http://www.mail-archive.com/nlpatumd@yahoogroups.com/msg00008.html</a>	3
3 (6%)	AltaVista	<a href="http://www.cs.toronto.edu/colloq/1999/broder.html">http://www.cs.toronto.edu/colloq/1999/broder.html</a>	4
2 (4%)	תשובה מלאה	<a href="http://www.computer.org/tab/tclist/tcmf.htm">http://www.computer.org/tab/tclist/tcmf.htm</a>	5
2 (4%)	AltaVista	<a href="http://www.infotoday.com/newsbreaks/nb000320-1.htm">http://www.infotoday.com/newsbreaks/nb000320-1.htm</a> + <a href="http://www.altavista.com/about">http://www.altavista.com/about</a>	6
2 (4%)	תפקיד	<a href="http://www.cc.gatech.edu/~lingliu/panels/WWWPanelMay13am.htm">http://www.cc.gatech.edu/~lingliu/panels/WWWPanelMay13am.htm</a>	7
2 (4%)	כתובת	<a href="http://www.ee.technion.ac.il/people/zivby/papers/decay/decay.pdf">http://www.ee.technion.ac.il/people/zivby/papers/decay/decay.pdf</a>	8

כתובת ה-URL הראשונה בטבלה 3 מספקת תשובה מלאה לשאלה. ה-URL השני נותן מידע מלא על תפקידו, אך חסר בו מידע על כתובת מקום עבודתו של אנרי ברודר. לדף זה הגיעו שמונה משתתפים, ויש לשער שעוד מתחרים היו מגיעים אליו לולא טעות הכתיב ב-URL (border במקום broder). בדף השלישי ברשימה ביקרו שישה מתחרים ושלושה מהם הסתפקו בתשובה החלקית ביותר: חוקר במרכז המדעי ע"ש Watson של IBM. בדף זה מופיעה הערה מעניינת:

Strangely enough, for a person who does research on the web, Andrei doesn't have a home page, or at least doesn't have one that can easily be found. Perhaps that should tell us something? (:)

אל ה-URL הרביעי ברשימה הגיעו שבעה משתתפים, רק שלושה הסתפקו במידע הנמצא בדף זה (עובד ב-AltaVista) אך גם מתוך אלה שהמשיכו את חיפושיהם, שניים נוספים הגיעו למסקנה שהוא עובד ב-AltaVista. המספר הממוצע של כתובות ה-URL שנבדקו עד לתשובה הסופית היה שמונה בממוצע (בהתבסס על 33 הלוגים התקינים). הממוצע עולה ל-9.2 אם אנחנו מתייחסים רק למשתתפים שהצליחו להשיב על השאלה במלואה. היו מקרים בהם נבדקו יותר מ-20 דפים ולעומת זאת היו גם מקרים בהם נבדקו 2-3 דפים בלבד. כל המשתתפים, פרט לשלושה, החלו את החיפוש ב-Google. שלושת היוצאים מהכלל חיפשו תחילה ב-Yahoo, ב-AltaVista וב-Answers.com. רוב המשתתפים היו נאמנים לאורך כל הדרך ל-Google, ורק מעטים ניסו את מזלם גם בכלי חיפוש אחר. קצת יותר משתתפים נעזרו בחיפוש מקומי באתרים (הם הגיעו לאתרים אלה באמצעות מנועי החיפוש הכלליים) - באתר של IBM וב-ACM Digital Library. כפי שראינו בטבלה 2, רוב המשתתפים החלו עם השאלתה המתבקשת מעליה: *Andrei Broder*. תוצאות חיפוש זה (לפחות ב-Google) לא עזרו במיוחד.

כיצד המשיכו המשתתפים בחיפושיהם? חלקם עיינו בתקצירי התשובות (snippets) בלבד ודלו משם מילות חיפוש נוספות, חלקם ביקרו בחלק מהתוצאות ו/או ערכו חיפושים מקומיים באתרים שאליהם הגיעו במטרה למצוא תשובה לשאלה או כדי לשנות את השאלתה לשאלתה מוצלחת יותר. ניתן לומר, שתהליכי איתור המידע של המשתתפים מתאימים למודל של (Bates, 1999) Marcia Bates הנקרא berrypicking model. במודל זה אוספים המשתמשים פיסות מידע (berrypicking) ומשנים את השאלתות שלהם על סמך המידע שנאסף במהלך החיפוש. במספר מקרים המשתתפים ציינו יותר מ-URL אחד עליהם מתבססת תשובתם הסופית, מה שמחזק את ההתאמה בין המודל למה שהתרחש בפועל במהלך התחרות. נדגים את תהליכי החיפוש של שני מתחרים: משתתף א' שהצליח להגיע לתשובה הנכונה ומשתתף ב' שהגיע למסקנה שברודר עובד בחברת AltaVista.

## הדגמת תהליכי החיפוש משתתף א

■ משתתף א' החל את החיפושים עם השאלתה המתבקשת *Andrei Broder*. השאלתה הופנתה ל-Google.



DBLP: Andrei Z. Broder, הוא ביקר בתוצאה הראשונה, ■

[http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/b/Broder:Andrei\\_Z=.html](http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/b/Broder:Andrei_Z=.html) אך לא הצליח למצוא בדף זה מידע רלוונטי.

הוא ניסה את מזלו עם השאלתה *Andrei Broder homepage* ב-Google. ■

בדף התוצאות של שאלתה זו הקליק על הכתובת <http://theory.lcs.mit.edu/~iandc/Authors/broderandreiz.htm>, אך גילה שהגם שהדף מתייחס לברודר ועל אף שהמילה homepage מופיעה בו, היא אינה קשורה לברודר, אלא מובילה לדף הבית של כתב העת Information and Computation.

משתתף א' החליט שוב לשנות את השאלתה והפעם חיפש ב-Google את המונחים *Andrei Broder address*. הפעם לא בחר באחת התוצאות והסתפק בעיון ב-snippets והחליט שכדאי להגביל את החיפוש לאתר של IBM.

החיפוש הבא היה *Andrei Broder site:ibm.com*. ■

התוצאה הראשונה עבור חיפוש זה היה הדף:

<http://domino.watson.ibm.com/arccs/csnotesx.nsf/f07e5510fad528f9882566890024a84f/ee2db79af3089.26688256fe30079609e?OpenDocument&Click=>

הדף הזה אינו נגיש יותר, אך בזמנו הכיל תקציר של הרצאה שניתנה על ידי אנדרי ברודר במרכז המדעי של IBM ב-Almaden באפריל 2005. מדף זה ניתן היה ללמוד שבאפריל 2005 היה ברודר מועסק ב-IBM (והיה זה מידע עדכני, כי התחרות נערכה במאי 2005). המשתתף החליט לבחון דף נוסף מרשימת התוצאות:

<http://www.almaden.ibm.com/cs/people/ravi/waw2003/waw2003-cfp.html>

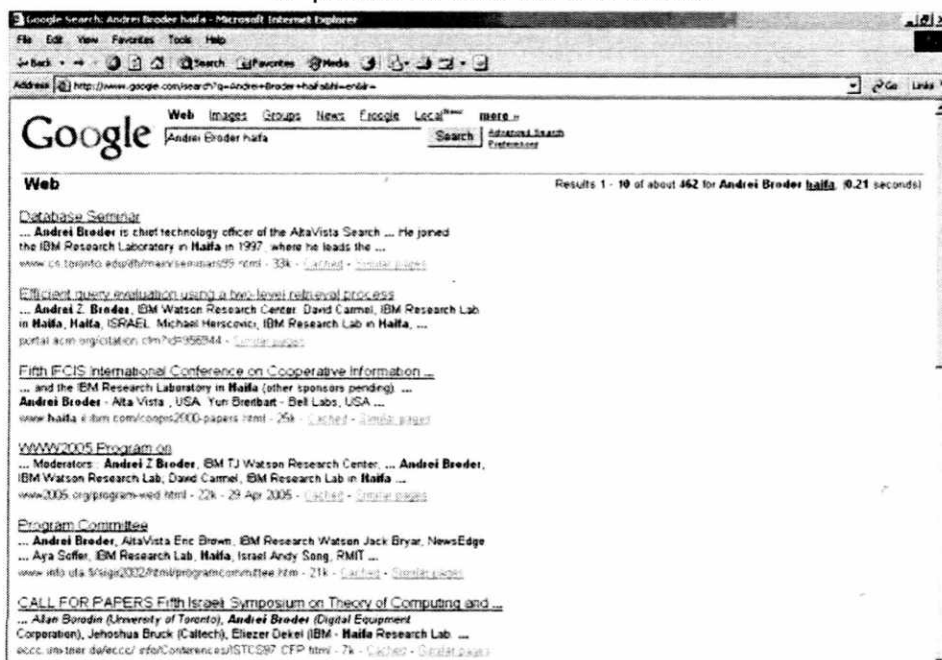
גם דף זה כבר אינו נגיש, אך ניתן לאתר אותו בארכיון האינטרנט, בכתובת:

<http://web.archive.org/web/20050305090306/http://www.almaden.ibm.com/cs/people/ravi/waw2003/waw2003-cfp.html>

דף זה לא הוסיף מידע, אך אישר שברודר עובד של IBM.

המשתתף החליט (אולי על סמך ה-snippets של הדף) המודיע על סמינרים באוניברסיטת טורונטו ■ <http://www.cs.toronto.edu/db/main/seminars99.html>, שברודר עובד ב-IBM Haifa ושינה את השאלתה ל-*Andrei Broder haifa*. היתה זאת החלטה שגויה, כי אמנם ברשימת ההרצאות הופיעה IBM Haifa, אך אזכור זה התייחס למרצה אחר בסדרת הסמינרים. למזלו של המשתתף, ברודר כתב מספר מאמרים עם חוקרים מ-IBM Haifa. בתרשים 8 מוצג צילום מסך של המשתתף.

## תרשים 8: צילום מסך של משתתף א'



■ המשתתף החליט לבחון אח התוצאה השנייה עבור השאילתה *Andrei Broder Haifa*, תקציר של מאמר שברודר כתב ביחד עם חוקרים מ-IBM Haifa. התקציר (<http://portal.acm.org/citation.cfm?id=956944>) היה חלק מה ACM Digital Library. הקלקה על שם של מחבר מעביר את המשתמש לרשימת המאמרים הממופתחים על ידי ACM Digital Library שונתבו על ידי אותו המחבר. אחת התוצאות הראשונות ברשימה היה מאמר בשם "Exploring, Modeling and Using the Web Graph" <http://delivery.acm.org/10.1145/870000/860436/p1-broder.pdf?key1=860436&key2=1300267411&coll=GUIDE&dl=GUIDE&CFID=75879203&CFTOKEN=96863623>

אילו היה מגיע למאמר עצמו, היה מגלה מה היה תפקידו של ברודר באותה העת, אך הוא לא המשיך בכיוון זה.

■ הוא חזר לדף תוצאות החיפוש והחליט לבדוק תוצאה נוספת, את הדף בכתובת <http://www2005.org/program-wed.html>, שהופיעה, בין התוצאות מכיוון ששותף שלו למאמר, דוד כרמל, עובד ב-IBM Haifa. בדף זה היה כתוב כי אנדרי ברודר עובד ב-IBM Watson Research Lab.

■ כנראה שבגלל המידע שראה בדפים הנ"ל, המשתתף השתכנע לשנות את השאילתה בפעם נוספת השאילתה החדשה היתה: *Andrei Broder IBM Watson Research Lab*. במקביל לשאילתה זו הריץ את השאילתה *Andrei Broder*, ב-Google Scholar אך כנראה שלא היה מרוצה מהתוצאות, כי לא המשיך בכיוון זה.

■ ה-snippet של אחת התוצאות

"Towards the next generation of enterprise search technology - Author Bios"

■ snippet- ב (<http://www.research.ibm.com/journal/sj/433/brodeaut.html>) נראה מביטיח במיוחד, כי כבר ב- הופיעה כתובת כלשהי. בחירה זו הוכיחה עצמה כמוצלחת (ראה תרשים 6), כי הדף הזה הכיל את התשובה לשני חלקי השאלה ( snippet הוא רצף המילים הקצר המופיע מתחת לכותרת של תוצאת החיפוש, לפעמים ניתן להסיק מתוכנו את מידת הרלוונטיות של תוצאת החיפוש).  
■ תהליך החיפוש לקח כ-15 דקות.

## משתתף ב'

■ גם משתתף זה החל את חיפושיו ב-Google בשאילתה *Andrei Broder*. גם הוא בחן את התוצאה הראשונה עבור שאילתה זו: DBLP: Andrei Z. Broder ([http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/b/Broder:Andrei\\_Z=.html](http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/b/Broder:Andrei_Z=.html)).

■ הוא החליט לבחון דף נוסף מרשימת התוצאות. דילג על הדף "No such user", אבל הקליק על דף בשם "www9 paper" <http://www.almaden.ibm.com/cs/k53/www9.final> - כיום ניתן למצוא אותו באמצעות ארכיון האינטרנט. זהו מאמר ידוע של אנדרי ברודר ושותפיו: "Graph structure of the Web". במאמר זה רשומה AltaVista Company, San Mateo, CA כמקום עבודתו של אנדרי ברודר.

■ למרות פיסת מידע זו, השאילתה הבאה של המשתתף היה *Andrei Broder IBM* (אולי על סמך אחד ה-snippets בדף התוצאות של השאילה הקודמת). הוא לא ביקר באף אחד מהדפים שהופיעו ברשימת התוצאות של השאילתה החדשה.

■ משתתף ב' שוב שינה את השאילתה, הפעם ל-*Andrei Broder AltaVista*. במקביל, ביקר גם באתר Yahoo! People Search, אך לא ביצע שם חיפוש.

■ מתוצאות החיפוש של השאילתה *Andrei Broder AltaVista*, בחר את הדף

"AltaVista Introduces Advanced-Search Center, New Resources for Power Searchers and Webmasters" <http://www.infotoday.com/newsbreaks/nb000320-1.htm>. המשתתף לא שם לב לכך שתאריך פרסום הכתבה שהופיע בדף היה 20 במרץ 2000, או שבחר להתעלם מהמידע הזה, אבל התייחס לקטע הבא: "Andrei Broder, vice president of research at AltaVista"

■ על דף זה הופיע קישור ל-AltaVista (<http://www.altavista.com>). הוא הקליק על הקישור והמשיך דרך קישור נוסף ל-About AltaVista (<http://www.altavista.com/about/>).

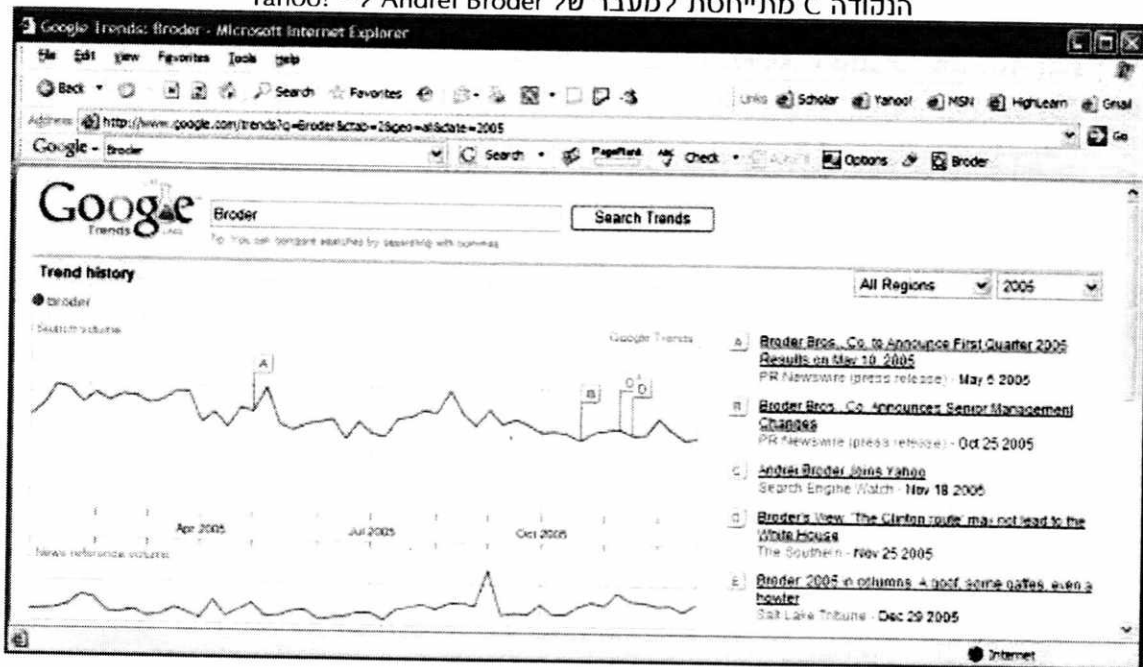
■ בדף זה הופיעה כתובת החברה: 74 North Pasadena Avenue, 3rd Floor, Pasadena, California 91103.

■ משתתף ב' הסתמך על שני דפי Web: <http://www.infotoday.com/newsbreaks/nb000320-1.htm> ו- <http://www.altavista.com/about/> ובתוך שש דקות הגיע למסקנה שאדרי ברודר הוא סגן נשיא לענייני מחקר ב-AltaVista אשר בפסדנה, קליפורניה.

## סיכום ומסקנות

ה-Web הוא מקור מידע עשיר מאוד, ניתן למצוא בו מידע כמעט על כל דבר, אך בגלל כמויות המידע העצומות לפעמים דומה החיפוש בו לחיפוש מחט בערימת שחת. חיפוש מידע עדכני על חוקר האינטרנט אנדרי ברודר הוא דוגמה מצוינת לכך. המידע הראשון שמתקבל עליו הוא שאין משתמש כזה ("no such user").

**תרשים 9:** חיפושים עבור Broder לאורך זמן, כפי שנמסר ע"י Google Trends  
הנקודה C מתייחסת למעבר של Andrei Broder ל- Yahoo!



בסוף שנת 2006 הרבה יותר קל למצוא מידע עליו מאשר במאי 2005 (מועד תחרות החיפוש באינטרנט). המעבר של ברודר ל- Yahoo! היה אירוע מתוקשר היטב (ראו תרשים 9). מאז יוני 2006 יש לאנדרי ברודר דף בית בחברת Yahoo! בכתובת <http://research.yahoo.com/~azbroder>. דף זה מאוחזר עבור חיפוש שנערך ב-1 בדצמבר 2006, אך הוא מופיע רק במקום ה-26 ברשימת התוצאות של Google, במקום ה-33 בתוצאות של Yahoo!, ובמקום ה-30 ברשימת התוצאות של Windows Live (MSN לשעבר). כלומר, דף זה טרם זכה לחשיפה גדולה. ביחד עם זאת, יש לציין, שצמד המילים Andrei ו-Broder מופיע במספר גדול מאוד של דפי Web; מספר הדפים לפי Google הוא 130,000. מעניין יהיה לראות מתי דף ה-"No such user" ייעלם סוף סוף ממאגרי המידע של מנועי החיפוש. במאמר זה ניתחנו את דרכי החיפוש של משתמשים מנוסים. אנו ממליצים על עריכת מחקרים נוספים בכיוון זה.

מקורות

Bar-Ilan, J (2002). How much information search engines disclose on the links to a Web page? – A longitudinal case study of the 'Cybermetrics' home page, *Journal of Information Science*, 28(6): 455-466.

Bates, M.(1989). The design of browsing and berrypicking techniques for the online search interface, *Online Review*, 13:07-424, at <http://www.gseis.ucla.edu/faculty/bates/berrypicking.html>, accessed 4 May 2006.

Broder, A. (2002). A taxonomy of Web search, *ACM SIGIR Forum*, 36(2) at <http://www.sigir.org/forum/F2002/broder.pdf>, accessed 4 May 2006.

comScore (2005). *comScore reports July 2005 search engine rankings*, at <http://www.comscore.com/press/release.asp?press=622>, accessed 4 May 2006.

Enquiro (2005). *Did-It, Enquiro and Eyetools uncover search's golden triangle*, at <http://www.enquiro.com/eye-tracking-pr.asp>, accessed 4 May 2006.

Forums.Searchenginewatch (2004). Google say not reporting all backlinks ), at <http://forums.searchenginewatch.com/showthread.php?t=2423&page=1&pp=20>, accessed 4 May 2006.

Yahoo! Media Relations (2005). *Web Expert and Search Pioneer to Apply Information Retrieval Expertise to Yahoo! Research Projects*, at <http://docs.yahoo.com/docs/pr/release1271.html>, accessed 4 May 2006.

Yale Information Society Project (2005). at <http://islandia.law.yale.edu/isp/regulatingsearch.html#broder>, accessed 4 May 2006.

www.ziur.co.il info@ziur.co.il 03-9249091 **זיעור**

ניהול המידע בארגון, שיווק מאגרי מידע

- הפצת מאגרי מידע בכל התחומים
- הפצת כחבי-עת בגרסה מודפסת, אלקטרונית וע"ג מיקרופילם
- שיווק עבודות דוקטורט וחזוה מכל העולם
- מיקרופילם וסריקה מסמכים

Engineering Village 2  
 HW Wilson  
 Swets Information Services  
 THOMSON MICROMEDEX  
 factiva  
 CHADWYCK-HEALEY  
 Safari  
 Knovel  
 Bowker  
 ProQuest  
 OVID SilverPlatter  
 ELSEVIER SCIENCE @ DIRECT

רח' מגשימים 20 ת.ד. 7657 קריית מטלון, פתח תקוה 49170