

סיכום מצב קיים במערכות אחזור טקסט

רות הנדל

אחזור טקסטואלי ממאגרי מידע - מצב נתון דר' רות הנדזל

מאגרי המידע העולמיים כוללים ספרות מקצועית בכל תחומי הפעילות האנושית: מתחומי המדע ועד לנתוני חברות, מנתונים על סרטים ועד נתונים דמוגרפיים של מדינות שונות.

המאגרים מיוצרים ע"י גופים שונים: מכוני מחקר, אוניברסיטאות וחברות פרטיות.

השמוש במאגרים נעשה באמצעות ספקי המידע, שהם חברות המחזיקות במחשביהם מספר רב של מאגרים. בין ספקי המידע הגדולים יש לציין את IRS-ISA, QUESTEL, STN, KRI (DIALOG DATASTAR) ורבים אחרים. המאגרים מתעדכנים בתדירות שונה: אחת ליום (עתונות יומית) מאגרים מדעיים (לרוב אחת לשבוע) ועוד.

ניתן לסווג את המאגרים למספר קבוצות:

1. מאגרים ביבליוגרפיים. קבוצה זו של מאגרים מכילה פריטים ביבליוגרפיים כלומר, הפניה למאמרים או דוחות מחקר שהתפרסמו עפ"ר בכתבי עת מדעיים. כל רשומה במאגר ביבליוגרפי כוללת מס' שדות: כותר, תמצית, מלות מפתח (כל אלו מצויים באינדקס הבסיסי), שם מחבר/ים, כתובת המוסד, שם כה"ע וכו' (אלו שדות נמצאים באינדקסים ספציפיים). במאגרים ביבליוגרפיים קיימת בעיה, שהטקסט המאוחסן במחשב מוגבל כפי שצוין לעיל. אי לכך הכרחי להשתמש במלון מבוקר, תזאורוס, כלומר, רשימת מושגים המאורגנים באופן היררכי. המפתח, לו חלק מתהליך יצור המאגר, מצרף לכל מסמך את מלות המפתח המתאימות לפי חוקי המערכת. מלות המפתח מבטאות את תוכן המסמך כולו ומסייעות למחפש המידע.
2. מאגרים שהם טקסט מלא. אלו מאגרים הכוללים את הטקסט המלא של המסמך כגון: עתונות יומית, פרוטוקלים רפואיים, אנציקלופדיות מקצועיות ועוד.
3. בנקים של נתונים. אלו הם מאגרים המכילים נתונים אלפא - נומריים, בתחומים שונים: פרופילים של חברות, רשימת תופעות לואי של תרופות ועוד.

שיטות האחזור מרכבות באופן בסיסי מחפוש בולאני וחפוש בטקסט. החפוש הבולאני מאפשר קישור לוגי בין מושגים משדות שונים. השמוש הוא רק בשלושת האופרטורים הבסיסיים של AND, OR, NOT. החפוש בשפה חופשית שונה מספק למישנהו, כאשר בשפה גמישה מדובר באחזור של מילים הנמצאות במרחק נתון אחת מהשניה עם אפשרות להגבלת החפוש לשדה הספציפי.

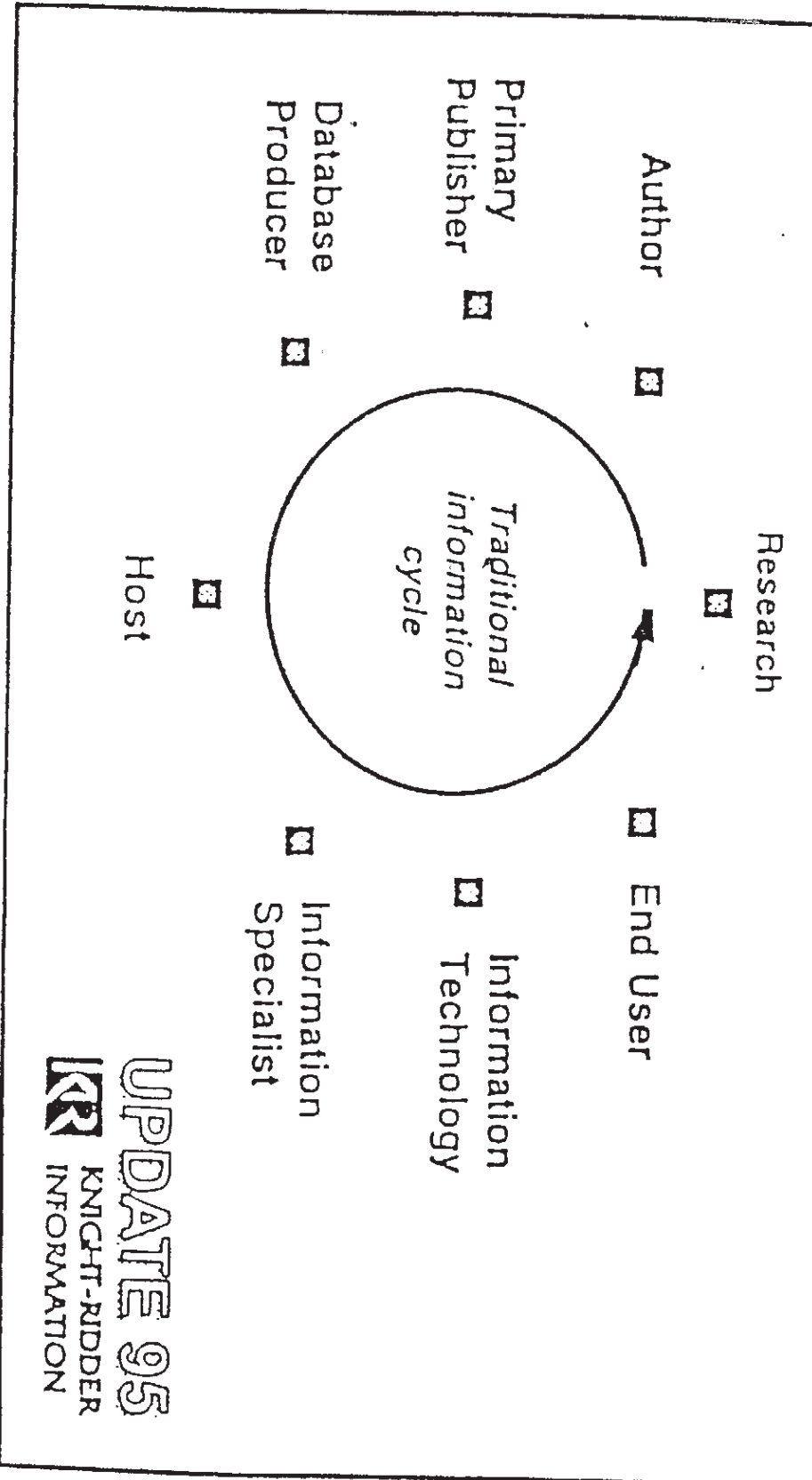
אופציות הגדום משמעותיות בכל אחזור טקסטואלי. הגדום בשפות הכתובות משמאל לימין הוא בסוף המילה, כאשר ניתן להגביל את הגדום למספר אותיות. יש שפות חפוש בהן קימת גם אפשרות לגדום באמצע המילה. תכונות אלו של גדום מציגות פתרון לבעיות הכתיב השונה כתוצאה מטרנסליטרציה, אחזור בטקסט המערב באנגלית ובאמריקאית ועוד...

עבודת אחזור המידע דורשת בנוסף להכרת המאגרים גם שליטה בשפות האחזור של הספקים.

כיום, מענינים יצרני המאגרים וספקי המידע להרחיב את מס' הצרכנים. לשם כך משקעת עבודת פתוח ליצירת ממשקים ידידותיים הנוחים לשמוש של המשתמש הלא מקצועי - משתמש הקצה.

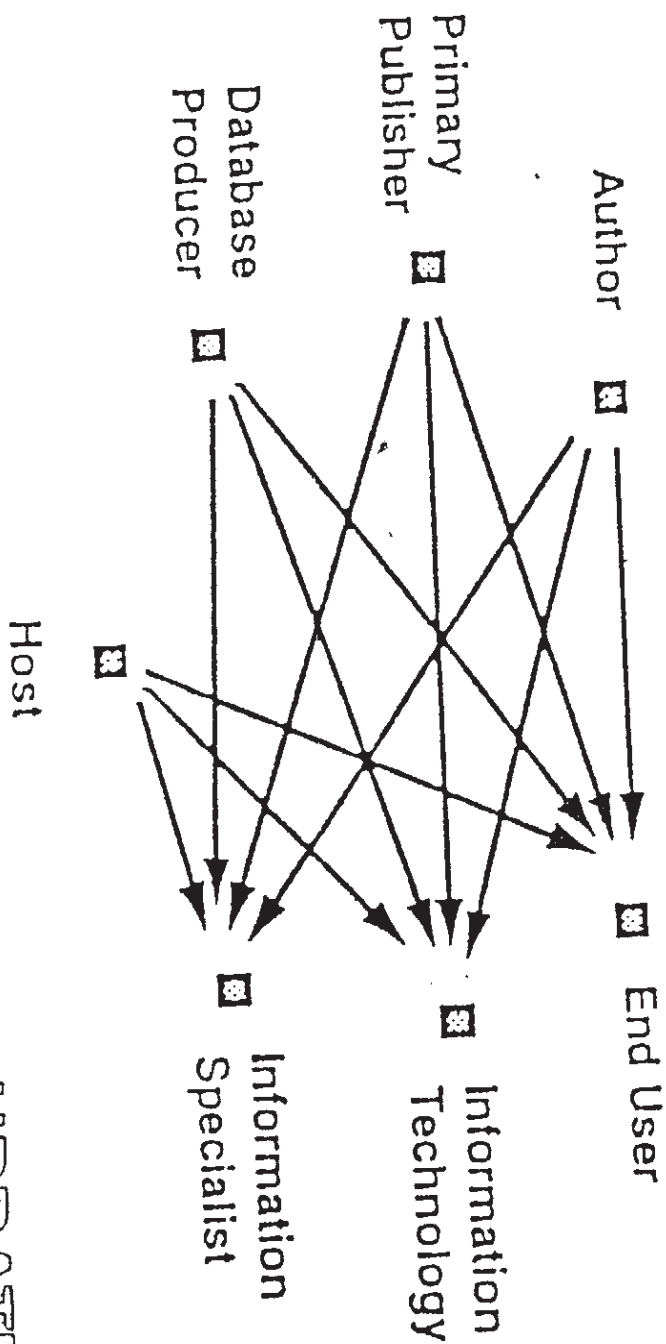
הממשקים הם מסוגים שונים: מסכי בחירה, רשתות סמנטיות המסיעות בבחירת המושגים במאגר ספציפי, כלי אחזור סטטיסטיים המתייחסים לשכיחות הופעת המושגים ועוד...

WE are on the Move...



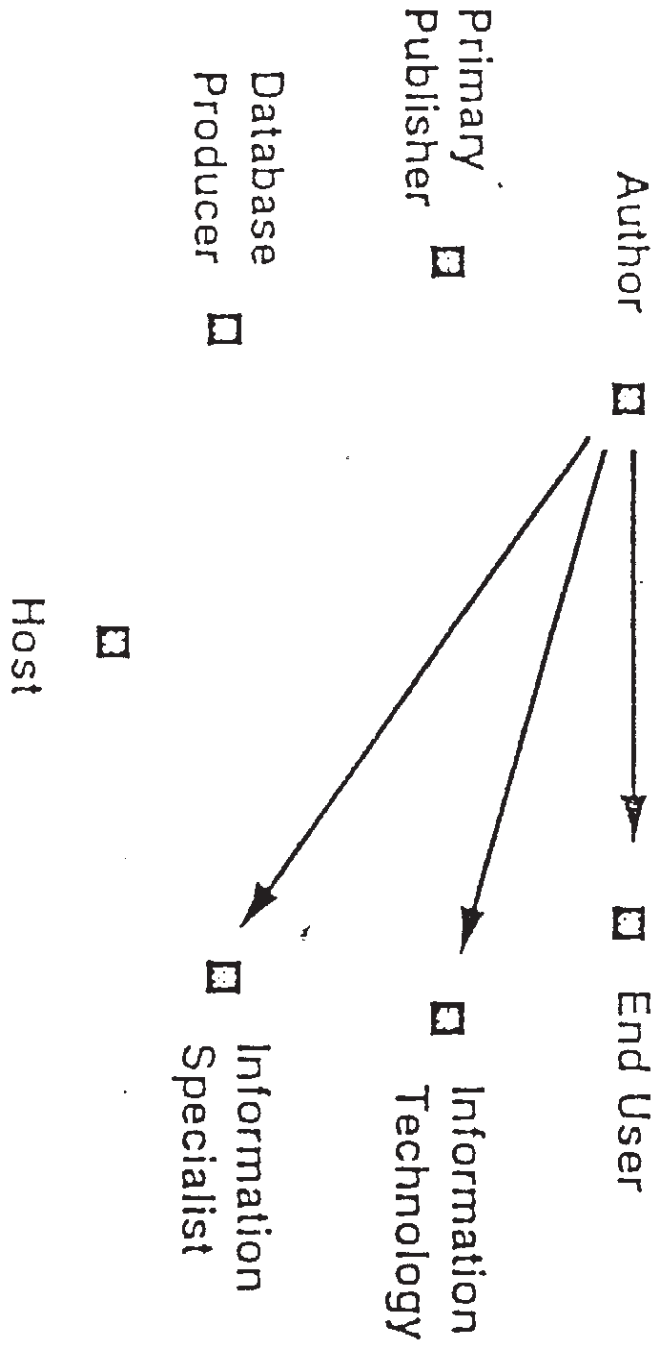
UPDATE 95
KNIGHT-RIDDER
INFORMATION

WE are on the Move...



UPDATE 95
KNIGHT-RIDDER
INFORMATION

WE are on the Move...



UPDATE 95
KNIGHT-RIDDER
INFORMATION