

בסיסי נתונים/מורן סרף

טור לשבוע: 11.2.2003 – 17.2.2003

כמות הנתונים שמחזיק המוח אינה ניתנת לכימות, אך ניתן להעריך כי היא במידה רבה מהווה פונקציה של כמות הניירונים במוח. הכמות הזאת קרובה לכמה מיליארדים של תאים. לכל אחד ממיליארד תאים אלו עשויים להיות עשרות קשרים סינפטיים לאחרים – וכך אנו עולים למספר גדול הרבה יותר של רכיבים המשתתפים בנייה הזיכרון. תחת עומס אינפורמציה זה אנו עדים למנגנון יעיל ביותר המבצע שליפות במהירויות מדהימות. נדיר שאנו נתקלים במקרה בו אנו מבקשים ממוחנו נתון זה או אחר ונתקלים בזמן קבלה שאינו מידי. אנו מתקשים לדמיין מצב בו נרצה להיזכר בשמו של אדם מסוים, ונדע בוודאות שתהליך ההיזכרות עתיד לקחת דקות מספר ('סליחה, איני מצליח להיזכר בשמך... תן לי רק עוד כמה דקות, אני בדיוק שולף את זה...!'). מאידך – אנו מכירים בכך שבסיסי הנתונים האורגני האמור עשוי לעיתים להכזיב כליל עת הוא פשוט יאבד מידע כזה או אחר....

מעניין לציין כי מערכת ה-caching של המוח דומה עד מאד למערכות הממוחשבות. אם בדיוק דיברנו על האדם שמולנו, או שזה מכבר עסקנו בו – נגלה כי הסיכויים שנזכור את שמו ועוד פרטים לגביו סבירים יותר.

באשר לאופן שמירת המידע - אנו מצפים שלכל נתון בבסיסי הנתונים תהיה מעין 'משקולת' שתציין את חשיבות זכירת המידע, ובהתאם תשמור אותו בצורה טובה יותר. יום הנישואין הוא נתון שאני מעריך שהמוח האנושי נוטה להדביק לו משקולת גבוהה בקרב חלק כזה או אחר של האוכלוסייה, לעומת זכירת שמותיהם של 17 חברי הכנסת לשעבר של סיעת ש"ס.

כיצד עובד, אם כן, הזיכרון במוח? התשובה היא שאיננו יודעים. הסברה השלטת נכון להיום היא כי הזיכרון אינו ממוקם במקום אחיד במוח, אלא מפוזר בכל חלקיו. הוא טמון, כפי שציניתי, בקשרים הסינפטיים בין הניירונים, והוא נתון לשינויים. מעבר לכך, איננו בטוחים בוודאות בדבר באשר אליו.

אנו יכולים לומר בוודאות שאופי הזיכרון הביולוגי שונה באופן מהותי מן הזיכרון הממוחשב – עד כי למעשה כל ניסיון השוואה בין המוח ובין מחשב מעורר שוב ושוב את זעמם של נירופיזיולוגים – הרואים בכך זילות של מחקרם. ברור כי בעוד שתמונה במחשב נשמרת כאשר כל הנתונים למעשה ארוזים כחבילה אחת של אינפורמציה במקום קבוע, אזי הנתונים במוח עוברים עיבוד וכיוון משמעותי. לו היינו זוכרים כל תמונה שראינו בחיינו בצורת ה-bitmap המתאים לה, במהירה היה מתמלא מוחנו בתמונות והיה מתפקד בערך כמו אותם 'בסיסי נתונים בקבצים' מלפני שלושים שנים... המידע ככל הנראה אגור בצורות אסוציאטיביות הרבה יותר, כשהאפשרויות לזליגה של מידע היא כמעט הכרחית.

בסיסי נתונים מהווים נדבך מהותי בתכנון רשת. איור הגליל המוכר מתרשימי הזרימה הפך בעשור האחרון לשימושי לא רק ככלי לשמירת נתונים, אלא כמערכת בפני עצמה לאינטגרציה של חלקים שונים מן המכלול. לשיתוף מידע בין תהליכים, ולתקשורת בין מודולי תוכנה העושים שימושים נרחבים ביכולות בסיסי הנתונים. המנגנונים נותרו דומים למדי לאילו המוכרים – טבלאות, שאילתות, עדכונים, SQL, LDAP, וכו... אבל לפתע נראה שהמערכת מהווה לא רק כלי לניהול רשומות בצורה פשוטה, כי אם גם כלי לתכנון, טיפול, ושימוש אחר בידע. בסיסי הנתונים אינו רק מיכל נוח לאחזור מידע, אלא דרך לעצב אופן עבודה מסודר ומתודולוגי. בסיסי נתונים כבר מכילים יכולות יתירות מורכבות, צורות שחזור מידע מתוחכמות, דרכים ל-caching של מידע, ועוד תכונות שכבר נראות כמובנות מאליהן – אך לא בהכרח היו מוכרות לאבותינו כשעבדו עם בסיסי הנתונים הישנים והמוכרים רק להיסטוריונים. למן הימים שהמונח בסיסי נתונים היה רק שם נרדף לקובץ גדול מאד עם נתונים מופרדים ברווחים ופונקציות תשאול פשוטות, דרך שמירת המידע בדפי זיכרון בעלי יכולת הפרדה לסגמנטים טבלאיים, ועד ליכולות המבוצרות של בסיסי הנתונים דהיום עברנו כברת דרך.

יתירה מזאת – נראה כי המידע שמחזיקים בסיסי הנתונים הללו נוטה להיות מידע מן הסוג הרגיש ובעל הערך הרב ביותר בארגון. מרבית המידע תיקים רפואיים, סיסמאות, אינוונטרים של ציוד צבאי סודי – כולם מוצאים את דרכם בסופו של היום – אל אחת מן הרשומות של בסיסי הנתונים, כזה צריך האחרון להיות מאד מוגן, והמידע בו צריך להיות מטופל ברגישות יתירה. שמור ומאובטח ("חיל", בסיסי הנתונים הוא ביתך. שמור עליו").

בהתאם לשבחים אלו שאני מרעיף על בסיסי הנתונים, ונכח כמויות המידע העצומות שנשמרות באלו האחרונים, ומשך זמני השליפה שעתידיים להפוך ארוכים עד מאד, אנו מצפים להתייעלות-רבתי באופן הניהול והשמירה של המידע. עיקר האלגוריתמים המשמשים את החברות הבונות את בסיסי הנתונים מתבססים על מחקרים מן האקדמיה שמוצאים דרכים כאלו ואחרות ליעיל את דרכי הגישה שלנו אל המידע – לרוב, תוך בניית אינדקסים חכמים יותר, ושיפור פונקציות השליפה.

בטור זה אני רוצה להפנות אצבע לעבר כיוון חשיבה אחר על אפשרות ייעול העבודה מול בסיסי הנתונים, כזה שינסה לדמות את אופן העבודה שבו פועל בסיסי הנתונים המוצלח ביותר המוכר לנו עד כה. כוונתי היא למוח האנושי.

כיצד נגיב כאשר חשבון הבנק שלנו יישכח כך פתאום... כשסתם כך ישכח בסיס הנתונים כי חולה זה או אחר רגיש לתרופה מסוימת... אולי כשנתייחס לא יפה אל בסיס הנתונים הוא יכעס עלינו ויעבד את המידע בצורה אחרת...

נראה, כי שימוש בבסיסי נתונים כאלו דורש, מעבר להבנה של אופן פעולת המוח, גם בחירה של סוג מסוים של מידע שאותו נסכים לשמור בו. בסיס נתונים כזה יכול להיות יעיל ביותר בשמירת מידע שרגישותו פחות קריטית לנו – דוגמת cookies של אתרים באינטרנט. במקרה הגרוע ביותר, כשהללו 'אבדו', נבקשם אותם שנית מן האתר. ומאידך, את שמירת המידע הרגיש שחשיבותו קריטית, נבצע עדיין בבסיסי הנתונים המוכרים מעולם המחשבים.

אולי בכלל הגיעה העת להתבונן על הקשר בין הביולוגיה לטכנולוגיה גם בכיוון ההפוך.

אולי הגיעה העת שנירופיזיולוגים ורופאים יחפשו דרכים ליעל את שמירת הנתונים במוחנו, באופן הדומה לנעשה בבסיסי הנתונים. אולי העולם יהיה טוב הרבה יותר כאשר תאריך יום הנישואין יוכל להיצרב אל מוחנו, בלא אפשרות לשכחו. הדבר בטוח ייטיב עם כמה זוגות נשואים בעולם...

במכון לבינה מלאכותית באוניברסיטת M.I.T עוסקים בשנה האחרונה בבניית מודל לבסיס נתונים מעין זה המוצע כאן. ייתכן והדור הבא של בסיסי הנתונים כבר בפתח.

כיום – אנו מסוגלים בעיקר 'לתת סימנים' בתכונות הזיכרון במוח, תוך איפיון של מעין קופסאות ומודולים של זיכרון. אנו יודעים כי קיים זיכרון קצר-מועד במוחנו – הוא האמון על זכירתם של פרטי מידע וגירויים מיידיים, ממנו עובר המידע – בעיקר נוכח חזרה שלנו עליו ועיסוק בו לזיכרון המכונה זיכרון העבודה, וממנו, בהתאם לקשרים הסינפטיים הנוצרים נוכח שימוש חוזר ונשנה באותו זיכרון עבודה, לעבר הזיכרון ארוך הטווח. מחקרים אחדים מציינים עוד תת-רכיבים נוספים בתרשים הזרימה הפשוט שציירתי זה עתה, ומציינים גם אפשרויות כאלו ואחרות של העתקת מידע מאזורים שונים לאחרים. כיוון שמטרתו של טור זה אינה לעסוק בחקר המוח, אסתפק בציון הללו כמאפיינים את זיכרונו – את בסיס הנתונים במוחנו.

ברוח הביולוגיה התופסת מקום מרכזי בהתפתחות הטכנולוגית של המאה העשרים ואחת, הגיעה השעה לנסות ולהרכיב בסיסי נתונים המחקים ומדמים במידה זו או אחרת את אופי עבודתו של המוח האנושי. בסיסי נתונים שישמרו את המידע בצורות אסוציאטיביות. שמירת הנתונים בצורה זו מאפשרת גישה מהירה אל המידע, ומוסיפה מעין משקלות למידע הנשמר.

גישה זו מבטיחה מצד אחד שליפות מהירות הרבה יותר, ועבודה יעילה בסדרי גודל – אך מכניסה לעולם המחשבים גם את בעיותיו של מוחנו...