

מבנית 9

שחר פרץ

7 במאי 2025

מרצה: טל קליגמן

נרחיב על B-Trees. אלו עצי חיפוש לא בינאריים, כאשר כל B-tree הוא בין d ל- $2d$ בניים. "הדרגה" – דרגת היציאה של הצומת. משום אופן לא מכיל מצביע להורה – תחזוק דבר כזה יקח המון פעולות IO (בגלל ספליטים). נרצה לבחור d בגודל של לא יותר מדף זכרון (סדר גודל של קילוביטים). חיפוש מבוצע ב- $O(\log_2 n)$ וכמות פעולות IO של $\log_d n$. הכנסה תיארך $d \log_d n$ עם $\log_d n$ פעולות IO.

במימוש bottom up קשה למקבל כמה פעולות בו זמנית, צריך למלא את המכסנית במסלול לצומת שמכניסים/מוחקים ועוד ("גלישה" זמנית של צמתים ממש לפני split ו-fuse). לעומת זאת up to bottom יכול לעשות fuse/splits מיותרים ולכן חסמי האמורטיזציה לא תקפים עליו. אם נרצה לקרוא את העץ לפי הסדר – in-order. מסתבר שבעבור מספר המפתחות k ו- L מספר העלים, מתקיים $L = k + 1$. נסמן ב- I את המפתחות הפנימיים. אז:

$$\text{מספר הילדים עץ} = \sum_{i \in I} K_i + 1 = \sum_{i \in I} K_i + \sum_{i \in I} 1 = K + I$$

או משהו כזה.

כדי להכניס רשימה מסודרת לעץ $(2, 4)$, נשתמש באלגוריתם הרקורסיבי "top-down" שבנה AVL בזמן לינארי. החסרון: יכול להיות שלא הכל יהיה באותו הגובה, עד לכדי 1. אפשר לתקן את זה בזמן לינארי באמצעות כך שנאחד את האחרון למעלה, אך עם זאת זה לא הפתרון המומלץ, בין היתר כי הוא לא עובד בעצים שאינם עצי $(2, 4)$. ישנו פתרון איטרטיבי יותר נורמלי שמפורט יותר גרפית במצגת.

עתה ננסה לעשות rank (כמו ב-AVLs) עם חסם I/O של $O(\log_d n)$. הפתרון הנאיבי הוא להשתמש בשדה size, אך הבעיה בו היא שהגישה ל- d הצמתים כדי למצוא את ה-size שלהם דורש $\Omega(d)$, וזה בלי לדבר על התחזוק. לכן נשמור את ה-sizes של הילדים בתוך הצומת עצמו. תחזוקת גדלים במצגת עצמה.

(הערה לעצמי: לעבור על שאלות 1, 4 בתרגול על B-trees)

מיונים

נדבר על Insertion, Selection, Bubble, Merge, Balanced-BST-sort, QuickSort ועוד. נחלק לשלושה סוגים: $O(n^2)$, $O(n \log n)$ ו- $O(n^2)$ עם expected של $O(n \log n)$. שאלות שאיכפת לנו מהן:

- | | | |
|---------------------------|------------------------------------|---|
| • האם זה in-place או לא? | • סיבוכיות זכרון | (מיון יציב) |
| • האם זה כבר ממיון חלקית? | • מספר פעולות ספציפיות, כמו I/O | • האם זה עובד עם streams (כל פעם לקבל חלק מהקלט)? |
| • סיבוכיות זמן | • מה קורה אם יש מספר שחוזר על עצמו | |

יש דרכים לייעל את insertion sort – לצורך הדוגמה, אם במקום לעשות הזזות נשמור את המספר שמסדרים בשבילו בצד (זה יחסוך פעולות IO אבל לא יעשה שיפור אסימפטוטי). מאוד יעיל לערכים נמוכים של n .

יש מבני נתונים שבהם אם יש דברים שחוזרים על עצמם נרצה לשמור על הסדר שלהם. נקרא לזה stable sorting (מיון שמשמר סדר איברים זהים).

.....

שחר פרץ, 2025

קומפל ב-L^AT_EX ונוצר באמצעות תוכנה חופשית בלבד