(a, b)-STROMY, STROMY S PRSTOM

©kuko

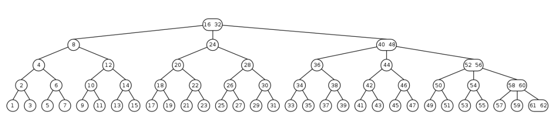
6.3.2018

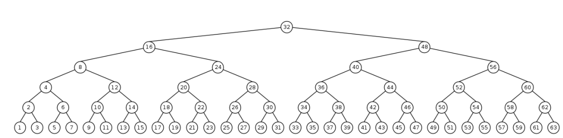
1 (a, b)-stromy

- (a, b)-strom pre $a \geq 2$, $b \geq 2a - 1$ je strom, kde
  - každý vrchol má $\leq b$ synov
  - každý vnútorný vrchol má $\geq a$ synov okrem koreňa, ktorý má ≥ 2 synov
  - všetky listy majú rovnakú hĺbku
  - vrchol s k kľúčmi $x_1, \ldots, x_k$ má $k + 1$ podstromov $T_0, \ldots, T_k$, pričom $T_0 \leq x_1 < T_1 \leq x_2 < \cdots < T_{k-1} \leq x_k < T_k$
- výška (a, b)-stromu je medzi $\Theta(\log_b n)$ a $\Theta(\log_a n)$
- vrchol s k synmi budeme volať k-vrchol
- napr. (2,3)-stromy, (2,4)-stromy; veľké a, b zvolené tak, aby sa jeden vrchol zmestil do 1 stránky pamäte sú vhodné pre externé dátové štruktúry, ktoré sú uložené na disku; pre $b = 2a - 1$ dostávame B-stromy
- jeden variant (pre $b = 2a - 1$ známy ako B^+-stromy) je uložiť všetky dáta do listov a vo vnútorných vrcholoch održovať iba kľúče, ktoré slúžia na navigáciu
- vyvažovanie, resp. udržovanie rovnakej hĺbky pre všetky listy dosahujeme nasledovne:
- pri vkladaní môžeme dostať vrchol s $b + 1$ synmi (b kľúčmi), ktorý je preplnený
  - (split) rozdelíme ho na 2 vrcholy s $\lceil (b+1)/2 \rceil$ a $\lfloor (b+1)/2 \rfloor$ synmi; rekurzívne potom musíme vyriešiť otca (ak je preplnený)
- pri vymazávaní môžeme dostať vrchol s $a - 1$ synmi, ktorý je príliš malý
  - (share) ak má suseda s $\geq a + 1$ synmi, môžeme mu najbližšieho krajného syna zobrať; dostaneme vrchol s a a $\geq a$ synmi a končíme
  - (join) ak má sused a synov, oba vrcholy spojíme; dostaneme jeden vrchol s $2a-1 \leq b$ synmi a rekurzívne musíme vyriešiť otca (ak je podtečený)
- počet rozdelení/spojení je v najhoršom prípade $O(\log_a n)$
- avšak ukážeme, že pre $b = 2a$ je rozdelení/spojení je len $O(1)$ amortizovane
- pre $b = (2 + \varepsilon)a$ sa dá dokázať odhad $\Theta(1/(\varepsilon a))$ amortizovane

2 $(2, 3)$-stromy

- ak do (2,3)-stromu vložíme prvky $1, 2, \ldots, 2^k-2$, dostaneme strom, kde všetky vrcholy obsahujú 1 kľúč, okrem vrcholov na pravej ceste, ktoré majú 2 kľúče
- ak následne do takéhoto stromu vložíme $2^k - 1$, všetky vrcholy na pravej ceste sa preplnia a musia sa rozdeliť ($\Theta(\log n)\times$split)
- naopak, ak potom prvok $2^k - 1$ odstránime, všetky vrcholy na pravej ceste sa zlúčia s bratmi a dostaneme pôvodný strom ($\Theta(\log n)\times$join)
- teda $m\times$ opakované vloženie a odstránenie $2^k - 1$ trvá $\Theta(m \log n)$ – (2,3)-strom nedokáže zaručiť $O(1)$ rozdelení/zlúčení ani amortizovane





- ak by sme však mali (2,3)-strom, do ktorého iba vkladáme a nikdy nemažeme, počet rozdelení bude $O(1)$ amortizovane

3 $(2,5)$-stromy

- v $(2,5)$-strome sa vrchol preplní pri 6 synoch
- split rozdelí 6-vrchol na dva 3-vrcholy ($6 \to 3+3$)
- join spojí 1-vrchol a 2-vrchol na 3-vrchol ($1+2 \to 3$)
- share prerozdelí 1-vrchol a $\geq$ 3-vrchol na dva $\geq$ 2-vrcholy ($1 + (\geq 3) \to 2 + (\geq 2)$)
- veta: počet rozdelení/spojení je $O(1)$ amortizovane
- invariant: každý 2-vrchol a 5-vrchol budú mať našetrený 1\$
- na insert si vyhradíme 2\$:
  - ak pri vkladaní 5-vrchol pretečie na 6-vrchol, vloženie aj rozdelenie zaplatíme z ušetrených peňazí
  - 1\$ použijeme na $O(1)$ operácií pri vkladaní do posledného vrcholu (ktorý už nepretečie), prípadne vytvorenie nového koreňa
  - 1\$ uložíme na účet, ak sme pritom vytvorili nový 5-vrchol
- na delete si vyhradíme 3\$:
  - ak pri vymazávaní 2-vrchol podtečie, vymazávanie aj spájanie zaplatíme z ušetrených peňazí
  - 1\$ použijeme na $O(1)$ čas na vymazanie v poslednom vrchole (ktorý už nepodtečie), prípadne operáciu share, prípadne zmazanie koreňa
  - po 1\$ vložíme na účet maximálne dvom novovytvoreným 2-vrcholom (posledný split alebo share)

- alternatívne, pri potenciálovej metóde by sme definovali $\Phi = \#2\text{-vrcholov} + \#5\text{-vrcholov}$
- ak sa pri vložení rozdelí k vrcholov, skutočná cena je $O(k)$, ale potenciál klesne o k, takže amortizovaná zložitosť je $O(1)$
- ak sa pri vymazávaní spojí k vrcholov, skutočná cena je $O(k)$, ale potenciál klesne o k, takže amortizovaná zložitosť je $O(1)$

4 $(2,4)$-stromy

- split: $5 \to 2+3$
- join: $1+2 \to 3$
- share: $1 + (\geq 3) \to 2 + (\geq 2)$
- dôkaz pre $(2,5)$-stromy na $(2,4)$-stromy nefunguje
  - pri $(2,5)$-stromoch boli „kritické“ 2- a 5-vrcholy a split/join vytvárali nové 3-vrcholy
  - pri $(2,4)$-stromoch sú „kritické“ 2- a 4-vrcholy a split vytvorí nový 2-vrchol, ktorý je kritický(!)
- na druhej strane, všimnime si, že join nevytvorí nový kritický vrchol a táto asymetria nás zachráni
- intuícia: 4-vrcholy budeme považovať za 2×horšie ako 2-vrcholy; tým pádom pri splite síce vytvoríme nový 2-vrchol, ale zbavíme sa 2×horšieho 4-vrcholu – celkovo sme si teda polepšili
- invariant: 2-vrcholy budú mať ušetrený 1\$, 4-vrcholy budú mať 2\$
- veta: počet rozdelení/spojení je $O(1)$ amortizovane
- na insert si vyhradíme 3\$:
  - ak pri vkladaní 4-vrchol pretečie, z ušetrených 2\$ zaplatíme 1\$ za rozdelenie ($O(1)$ čas) a 1\$ vložíme na účet novovytvorenému 2-vrcholu
  - 1\$ použijeme na $O(1)$ operácií pri vkladaní do posledného vrcholu (ktorý už nepretečie), prípadne vytvorenie nového koreňa
  - 2\$ uložíme na účet, ak sme pritom vytvorili nový 4-vrchol
- na delete si vyhradíme 3\$, argument je rovnaký ako pri $(2,5)$-strome

- alternatívne, pri potenciálovej metóde by sme definovali $\Phi = \#2\text{-vrcholov} + 2 \times \#5\text{-vrcholov}$
- ak sa pri vložení rozdelí k vrcholov, skutočná cena je $O(k)$; počet 2-vrcholov stúpne o k, ale #5-vrcholov klesne o k, takže celkovo Φ klesne o k a amortizovaná zložitosť je $O(1)$

- ak sa pri vymazávaní spojí k vrcholov, skutočná cena je $O(k)$, ale potenciál klesne o k, takže amortizovaná zložitosť je $O(1)$

5 Stromy s prstom