4. cvičení

Datové struktury I, 21. 10. 2024 https://iuuk.mff.cuni.cz/~chmel/2425/ds1/

Úloha 1 (Splay stromy mají potenciál)
Ujistěte se, že chápete, jak je definovaný potenciál ve splay stromech a že rozumíte hlavním myšlenkám analýzy amortizované složitosti splaye.

a) Jak je definován potenciál splay stromu?

b) Jaká je amortizovaná cena rotace (dvojité a jednoduché)?

c) Jaká je amortizovaná cena celého splaye a jak plyne z amortizovaných cen rotace?

d) Jaké je *reálná* cena celého splaye (a v jakých jednotkách ji vlastně počítáme)?

e) Jaký je potenciál perfektně vyváženého stromu? (Stačí nám rozumný horní a dolní odhad, pro jednoduchost předpokládejte $n = 2^k - 1$.)

f) Jaký je potenciál cesty? (Opět stačí rozumné odhady.)

Úloha 2 (Subset theorem)
Ukážeme následující tvrzení: mějme splay strom na vrcholech $[n]$ a podmnožinu $S \subseteq [n], s := |S|$. Pak m dotazů, které se ptají pouze na prvky S, má časovou složitost $\mathcal{O}(n \cdot s + m \log s)$.
Jako bonus jej můžeme zlepšit na $\mathcal{O}(n \cdot \min(s, \log n) + m \log s)$.
Předpokládejme, že máme posloupnost dotazů $q_1, \ldots, q_m$. Dále si označíme jako $T_0, \ldots, T_m$ stavy splay stromu, kde T_i je stav po i-té operaci, a jako $S_i := \{q_j : j \leq i\}, s_i := |S_i|$ si označíme množinu prvků, které byly dotazovány do i-té operace, a mohutnost této množiny. Budeme postupovat následovně:

a) Ukážeme, že když se podíváme na nejmenší podstrom T_i, který obsahuje všechny prvky S_i, tak tento podstrom obsahuje nejvýše $2s_i - 1$ vrcholů. Tento podstrom si dále označíme jako D_i. (D_0 je prázdný strom.)

b) Ukážeme, že to, co se děje při rotacích mezi T_i a T_{i+1} v D_i umíme simulovat v malém stromě, který odpovídá D_i.

c) Tím pádem můžeme cenu každého dotazu rozdělit na dvě části: cena na „zvednutí“ do D_i a cena uvnitř D_i. Triviálně odhadněte cenu na zvednutí s prvků jako $\mathcal{O}(n \cdot s)$.

d) Odhadneme za pomoci neváženě analýzy cenu všech operací uvnitř $D = (D_0, \ldots, D_m)$ (z prvního bodu máme odhad na velikost D_m) a uvědomíme si, že to stačí.

Bonus: Využijme věty ze cvičení a uvědomme si, že pro velké s ji také můžeme použít se stejným výsledkem.

Bonus 2: Jak to vypadá s částí a) pro naivní splay?

Úloha 3 (Semisplay)
V krocích LL a PP (neboli zig-zig) můžeme být línější a místo provedení dvojrotace provést jen jednoduchou rotaci vrchní hrany. Konkrétně v případě LL mějme tři vrcholy x, y a z takové, že x je levým synem y a y je levým synem z. Pak při splayování x provedeme jednoduchou rotaci hrany zy a dále pokračujeme ve splayování vrcholu y (místo x); původně splayovaný prvek se tedy *nemusí dostat* do kořene. Kroky LP a PL (zig-zag) nebo závěrečná jednoduchá rotace fungují stejně. Zanalyzujte amortizovaná cenu operace semisplay.

Úloha 4 (Ultralíné splay stromy)
Motivováni úspěšnou leností, která vedla k operaci semisplay, chceme na stromě provádět ještě méně změn. Co kdybychom při splayování prvku x rotovali jen každou druhou hranu na cestě z x do kořene a postupně měnili splayovaný prvek?

Úloha 5 (Rozděl a spojuj)
Pro splay strom T a hodnotu k navrhněte operaci SPLIT, která strom T rozdělí na dva stromy T', T'', přičemž ve stromě T' jsou všechny hodnoty menší nebo rovny k a ve stromě T'' jsou všechny hodnoty větší než k. Pokuste se zachovat amortizovanou složitost.
Dále pro splay stromy T', T'', kde všechny hodnoty v T' jsou menší než všechny hodnoty v T'', navrhněte operaci JOIN(T', T''), která sloučí stromy do jednoho v čase $\mathcal{O}(|T'| + |T''|)$.

Bonusové úlohy

Úloha 6 (Potenciál pro následníka)
Na prvním cvičení jsme si ukázali, že použití n operací následníka na libovolném BVS, když začneme ve vrcholu s nejmenším klíčem, má složitost $\mathcal{O}(n)$. Jak to můžeme dokázat pomocí potenciálu?

Zadání úkolu:

- ve každém testu stačí ~1 odstavec. Chtějí, abych to formuloval jako bakalářku!

Věta: Seq. Access Thm.

Pro splay strom na $\{1 \ldots n\}$: cena $(\text{Find}(1), \ldots, \text{Find}(n)) \in \mathcal{O}(n)$

Věta: Working set thm.

Strom T na $[n]$, $x_1 \ldots x_m \in [n]$ vyhledávání.

„relativně nedávno vyhledávané prvky mají relativně malou cenu"

↳ aka. nestihly spadnout moc hluboko

Označme z_i jako počet provedených různých findů mezi aktuálním a předchozím find(x_i).

Pak cena $(\text{Find}(x_1), \ldots, \text{Find}(x_m)) \in \mathcal{O}\left(n \cdot \log n + m + \sum_i \log(1+z_i)\right)$

Věta: Static finger thm.

T na $[n]$, $F \in [n]$, $x_1 \ldots x_m \in [n]$:

$$\text{cena}(\text{Find}(x_i)) \in \mathcal{O}\Big(\underbrace{n \log n}_{\text{cena potenciálu}} + \underbrace{m}_{\text{cena kroků}} + \underbrace{\sum_i \log(1+|x_i - F|)}_{\text{„čím blíže jsem prstu, tím levnější"}}\Big)$$

Úloha 1 (Splay stromy mají potenciál)
Ujistěte se, že chápete, jak je definovaný potenciál ve splay stromech a že rozumíte hlavním myšlenkám analýzy amortizované složitosti splaye.

a) Jak je definován potenciál splay stromu?

b) Jaká je amortizovaná cena rotace (dvojité a jednoduché)?

c) Jaká je amortizovaná cena celého splaye a jak plyne z amortizovaných cen rotace?

d) Jaké je *reálná* cena celého splaye (a v jakých jednotkách ji vlastně počítáme)?

e) Jaký je potenciál perfektně vyváženého stromu? (Stačí nám rozumný horní a dolní odhad, pro jednoduchost předpokládejte $n = 2^k - 1$.)

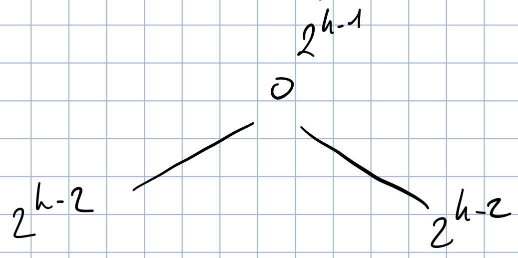
f) Jaký je potenciál cesty? (Opět stačí rozumné odhady.)

a) pro vrchol x: $\Phi(x) = \log_2(s(x))$

$$\Phi(T) = \sum_{x \in T} \phi(x)$$

b) $3 \cdot (\Phi(x') - \Phi(x))$ pro zigzag.., $1 + 3 \cdot (\Phi(x') - \Phi(x))$ pro zig

c) jeden splay $O(\log n)$



d) počet provedených rotací

e) $n = 2^{k-1}$

$$\leq \Phi \leq \sum_{i=0}^{d} (i+1) \cdot \log\left(2^{k-1-i}\right)$$

Úloha 2 (Subset theorem)

Ukážeme následující tvrzení: mějme splay strom na vrcholech $[n]$ a podmnožinu $S \subseteq [n], s := |S|$. Pak m dotazů, které se ptají pouze na prvky S, má časovou složitost $\mathcal{O}(n \cdot s + m \log s)$.

Jako bonus jej můžeme zlepšit na $\mathcal{O}(n \cdot \min(s, \log n) + m \log s)$.

Předpokládejme, že máme posloupnost dotazů $q_1, \dots, q_m$. Dále si označíme jako $T_0, \dots, T_m$ stavy splay stromu, kde T_i je stav po i-té operaci, a jako $S_i := \{q_j : j \leq i\}, s_i := |S_i|$ si označíme množinu prvků, které byly dotazovány do i-té operace, a mohutnost této množiny. Budeme postupovat následovně:

a) Ukážeme, že když se podíváme na nejmenší podstrom T_i, který obsahuje všechny prvky S_i, tak tento podstrom obsahuje nejvýše $2s_i - 1$ vrcholů. Tento podstrom si dále označíme jako D_i. (D_0 je prázdný strom.)

b) Ukážeme, že to, co se děje při rotacích mezi T_i a T_{i+1} v D_i umíme simulovat v malém stromě, který odpovídá D_i.

c) Tím pádem můžeme cenu každého dotazu rozdělit na dvě části: cena na „zvednutí" do D_i a cena uvnitř D_i. Triviálně odhadněte cenu na zvednutí s prvků jako $\mathcal{O}(n \cdot s)$.

d) Odhadneme za pomoci nevážené analýzy cenu všech operací uvnitř $D = (D_0, \dots, D_m)$ (z prvního bodu máme odhad na velikost D_m) a uvědomíme si, že to stačí.

Bonus: Využijme věty ze cvičení a uvědomme si, že pro velké s ji také můžeme použít se stejným výsledkem.

protože m operací v podstromu o s prvcích stojí:

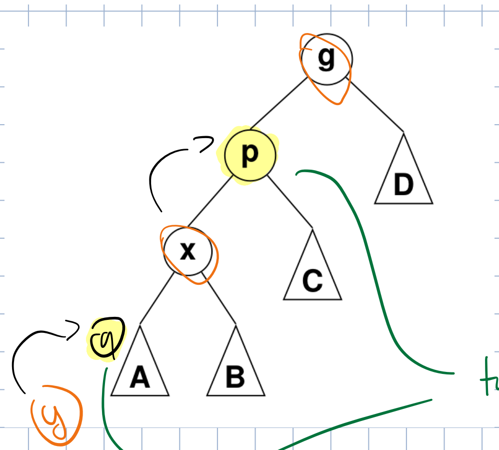
$O(s \log s + m \cdot \log s)$

$O(n \cdot s + m \log s)$

a řekneme, že probublat prvek od listu do mého podstromu trvá $O(n)$

to je triviální fakt, ale stačí nám to

a)



v nejhorším se mi tam mezi vyhledávanými prvky připlete vždy jeden nevyhledávaný prvek, který ale splay udrží nahoře

tyhle se mi tam jen „připletly"