Martin Hric
ID:111696

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLÓGIÍ

DÁTOVÉ ŠTRUKTÚRY A ALGORITMY
Vyhľadávanie v dynamických množinách

Vypracoval: Martin Hric

2020/2021

Martin Hric
ID:111696

ÚVOD

V tomto zadaní bolo našou úlohou implementovať vkladanie a vyhľadávanie v binárnych stromoch ako aj v hashovacích tabuľkách.

Ja som implementoval vlastný AVL strom, prevzal som Red and Black strom a prevzal som aj hashovaciu tabulku s chainingom, vlastnú hashovaciu tabuľku som neimplementoval.

Testovanie som robil na 10 miliónoch prvkov, keďže s viac prvkami som mal problémy kvôli nedostatku miesta v úložisku.

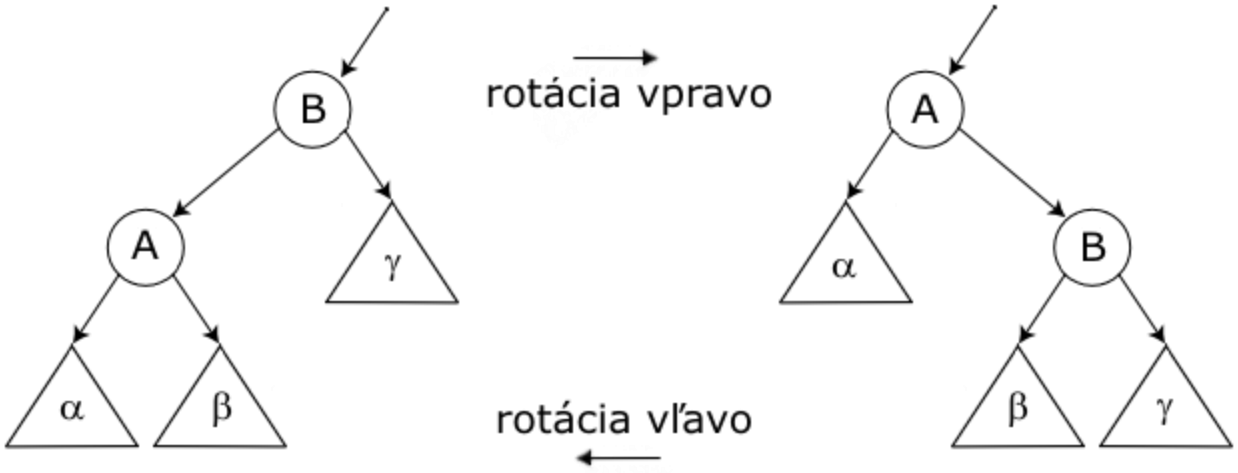
Martin Hric
ID:111696

AVL strom

V AVL strome sa pre každý uzol rozdiel výšky dvoch podstromov detských uzlov líšia najviac o jednotku, preto je známy aj ako výškovo vyvážený. Hľadanie, vkladanie, a mazanie majú zložitosť O(log *n*) v priemernom aj najhoršom prípade. Pridávanie a mazanie môže vyžadovať vyváženie stromu jednou alebo viacerými rotáciami stromu.

Existujú 4 druhy rotácii:

- Rotácia vpravo
- Rotácia vľavo
- Rotácia vpravo-vľavo
- Rotácia vľavo-vpravo



Ja som implementoval len 2 rotácie, ďalšie 2 som spravil ich kombináciou.

Martin Hric
ID:111696

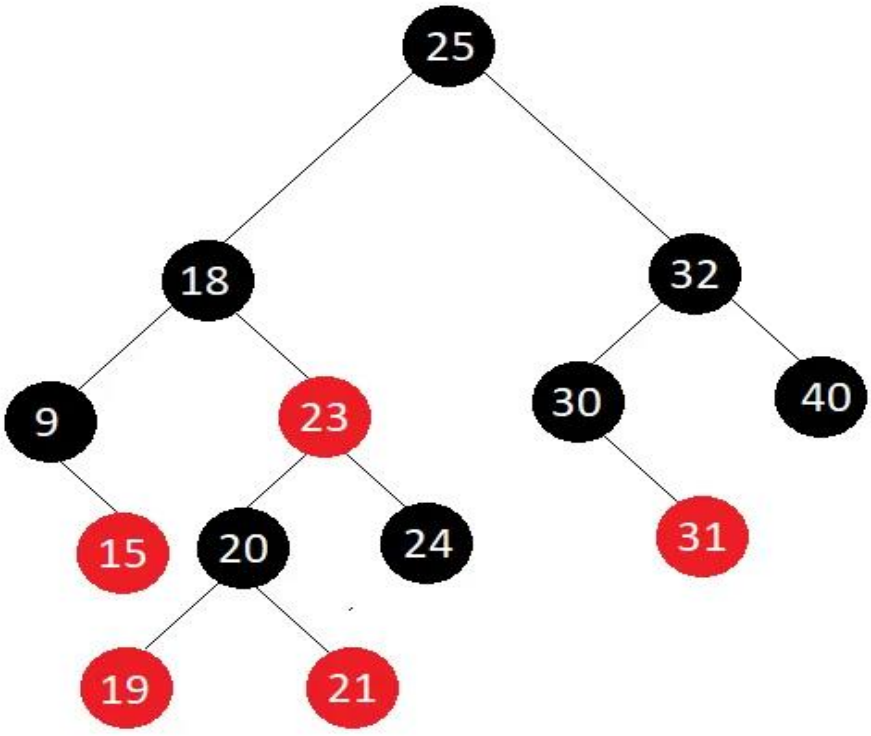
Red Black Tree:

ZDROJ: https://www.programiz.com/dsa/red-black-tree

Red Black Tree je samovyvažujúci binárny strom. V tomto strome je každý uzol červený alebo čierny. V tomto úvode červeno-čierneho stromu sa pokúsime pokryť všetky jeho základné vlastnosti.

Vlastnosti červeno-čierneho stromu

1. Každý uzol je zobrazený ako červený alebo čierny.
2. Koreňový uzol je vždy čierny.
3. Červený uzol nemôže mať červeného rodiča alebo červené dieťa (tj. Žiadne dva červené uzly nesusedia).
4. Každá cesta z ktoréhokoľvek uzla k jeho potomkovi NIL musí obsahovať rovnaký počet čiernych uzlov.



Valid Red-Black Tree

Martin Hric
ID:111696

Hashovacia tabuľka s reťazením:

ZDROJ: https://www.log2base2.com/algorithms/searching/open-hashing.html

V najjednoduchšej technike zreťazenej hašovacej tabuľky každé miesto v poli odkazuje na zreťazený zoznam vložených záznamov, pre ktoré hašovanie kľúča koliduje na rovnakú hodnotu. Pre vloženie je potrebné nájsť správne miesto a pridať ho na jeden koniec; pre mazanie je potrebné prehľadávanie zoznamu a odstránenie.

Zreťazené hašovacie tabuľky majú výhodu v porovnaní s tabuľkami s otvorenou adresáciou, že operácia odstránenia je jednoduchá a zmenu veľkosti tabuľky je možné odkladať oveľa dlhšie, pretože výkon sa znižuje oveľa pomalšie aj ak je miesto obsadené. Vlastne mnohé zreťazené hašovacie tabuľky nepotrebujú zmenu veľkosti vôbec, pretože degradácia výkonu je lineárna so zapĺňaním tabuľky. Napríklad zreťazená hašovacia tabuľka obsahujúca dvojnásobok odporúčanej dátovej kapacity by bola priemerne iba asi dvakrát pomalšia ako rovnaká tabuľka s odporúčanou kapacitou (za podmienky dobrej distribúcie kľúčov).

Zreťazené hašovacie tabuľky dedia nevýhody spojových zoznamov. Ak sa ukladajú malé záznamy, režijné náklady spojového zoznamu môžu byť významné. Ďalšou nevýhodou je, že prechádzanie spojovým zoznamom má nízky cache výkon.

Martin Hric
ID:111696

TESTOVANIE:

Testovanie som vykonával na 10 miliónoch prvkov , každé som spustil 3x a výsledky testovania sú vypísané v tabuľke, ktoré som potom spriemeroval.

Testoval som prvky od 0-9 999 999, zaradom, a potom náhodné čísla, ktorých bolo spolu 10 miliónov, v rozsahu od 1- INT MAX.

Veľkosť tabuľky bola nastavená na 1 milión, nemá resize.

SEARCH:

Lineárne testovanie 0 -9 999 999 prvkov			
AVL TREE		R&B TREE	HASH CHAINING
1.	1.262s	0.907001s	0.172001s
2.	1.187s	0.89s	0.172001s
3.	1.203s	0.908s	0.172001s
AVG:	1,2173s	0,901667s	0.172001s

Súčet random prvkov= 10 miliónov

Náhodné testovanie random prvkov			
AVL TREE		R&B TREE	HASH CHAINING
1.	10.231s	8.981s	17.039s
2.	9.275s	8.614s	15.897s
3.	9.258s	8.804s	16.824s
AVG:	9,588s	8,7996s	16,5867s

Martin Hric
ID:111696

INSERT:

Lineárne testovanie 0 -9 999 999 prvkov			
AVL TREE		R&B TREE	HASH CHAINING
1.	4.189s	3.409s	0.688s
2.	4.05s	3.409s	0.671999s
3.	4.025s	3.463s	0.705s
AVG:	4,088s	3,427s	0,68833s

Súčet random prvkov= 10 miliónov

Náhodné testovanie random prvkov			
AVL TREE		R&B TREE	HASH CHAINING
1.	18.896s	10.533s	18.801s
2.	16.814s	10.28s	17.429s
3.	16.602s	10.36s	18.107s
AVG:	17,437s	10,391s	18,112s

Martin Hric
ID:111696

ZÁVER

Lineárne testovanie:

INSERT trochu rýchlejší pri červeno čiernom strome oproti AVL.

SEARCH v podstate podobný. Výsledky môžu byť také, kvôli možno nie úplne optimálnej mojej implementácii AVL stromu.

Hashovacia tabuľka bola omnoho rýchlejšia oproti stromom, aj keď veľkosť tabuľky bola nastavená iba na 1 milión.

Random testovanie:

INSERT bol značne rýchlejší pri červeno čiernom strome oproti AVL.

SEARCH je opäť podobný pri stromoch. Časy narástli markantne pri random číslach, čo môže byť spôsobené viac rotáciami, ako aj vyššou výškou stromov.

Hashovacia tabuľka mala dokonca horšie časy ako mali stromy, ale to je kvôli veľkosti tabuľky, že je pomerne ale naschvál malá, teda vzniká vela kolízií.

Keď som tabuľku zvačšil na 2 milióny, časy rapídne klesly.