

B-stromy a jejich varianty

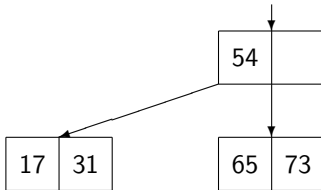
Michal Žemlička

1 B-stromy

B-strom (řádu m) je rozvětvený výškově vyvážený m -ární strom splňující tato omezení:

1. kořen má nejméně dva potomky a nejvýše m potomků, není-li listem;
2. každý uzel kromě kořene a listů má nejméně $\lceil m/2 \rceil$ a nejvýše m potomků;
3. každý uzel má nejméně $\lceil m/2 \rceil - 1$ a nejvíce $m - 1$ datových záznamů; kořen má nejméně 1 datový záznam;
4. všechny větve jsou stejně dlouhé

Popíšeme si dvě varianty B-stromů: neredundantní a redundantní. V prvním případě jsou data a klíče vždy pohromadě, ve druhém jsou data jen na listové úrovni a na zbývajících úrovních jen klíče v roli pomocných údajů usnadňujících hledání dat.



Obrázek 1: Ukázka B-stromu

2 Neredundantní B-stromy

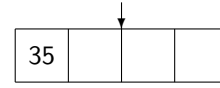
Na uzly neredundantních B-stromů můžeme pohlížet jako na záznamy se strukturou

$$[p_0, (k_1, p_1, d_1), (k_2, p_2, d_2), \dots, (k_n, p_n, d_n), u]$$

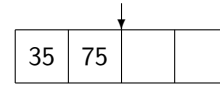
kde p_i jsou ukazatelé na potomky, k_i klíče, d_i data a u případný nevyužitý prostor (zarovnání na vhodnou velikost). Záznamy (k_i, p_i, d_i) jsou v uzlu uspořádány vzestupně dle klíčů.

Příklad 1: Vložení čísel 35, 75, 23 a 60 do prázdného (neredundantního) B-stromu s $m = 5$. Hodnoty se do uzlu vkládají tak, aby bylo zachováno uspořádání – viz jednotlivé části obr. 2.

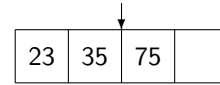
Příklad 2: Vložení čísla 28 do neredundantního B-stromu z příkladu 1 – viz postup zachycený obr. 3. Na počátku je uzel zcela zaplněn (obr. 3a)). Přidáním dalšího prvku (část b)) uzel přeteče. Obsah přeteklého uzlu si můžeme v pomocném prostoru rozdělit na tři části:



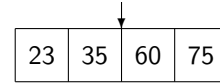
a) Vložení 35



b) Vložení 75 do stromu s 35



c) Vložení 23 do stromu s 35 a 75



d) Vložení 60 do stromu s 35, 75 a 23

Obrázek 2: Vkládání prvních prvků

1. na prvky, co zůstanou v původním uzlu
2. prvek, který o úroveň výše sehraje roli dělicí hodnoty
3. prvky, které přesuneme do nového uzlu

Části 1 a 3 by měly být stejně velké (je-li to možné). Pro prvek z části 2 zatím není vhodné místo. Vytvoříme proto nový uzel (nový kořen), kam tento prvek vložíme a kam umístíme odkazy na oba zbývajících prvky – viz obr. 3d).

Příklad 3: Do stromu z příkladu 2 vložíme ještě prvky 11, 15 a 31. První dva z těchto prvků půjdou vložit do levého listu (obr. 4). Při vkládání prvku 31 dojde k přetečení levého listu. Podobně jako v příkladu 2 obsah uzlu rozdělíme na tři části – obr. 5a).

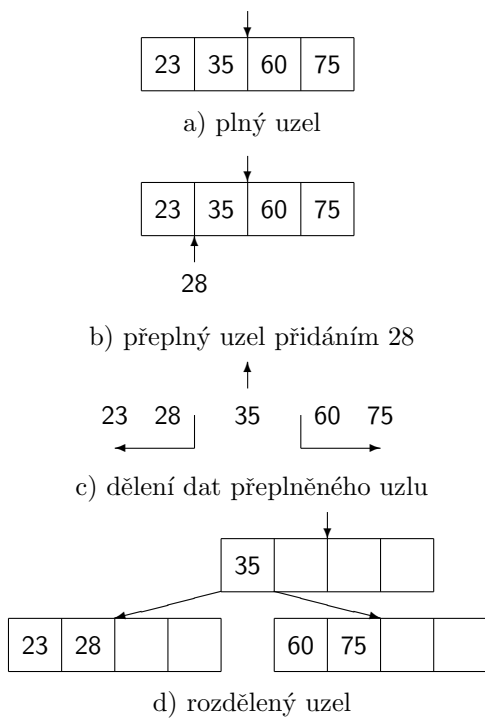
Příklad 4: Postupným přidáváním dalších prvků do stromu z obr. 5c) dostaneme strom se třemi úrovněmi, jako je na obr. 6.

Vypouštění

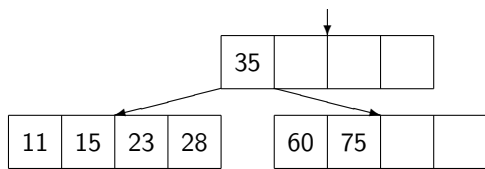
Z B-stromů je možné i vypouštět. I v tomto případě začneme vypouštěním z neredundantních B-stromů a až potom si ukážeme, jak je tomu u B-stromů redundantních.

Nejdříve se pokusíme prvek najít. Pokud jej ve stromě nenalezneme, není co vypouštět a vypouštění končí. Nelezneme-li prvek v listu, vypustíme jej.

Je ale třeba ověřit, zda nebyla porušena dolní mez naplnění uzlu. Pokud byla, je třeba situaci korigovat:



Obrázek 3: Štěpení uzlu B-stromu



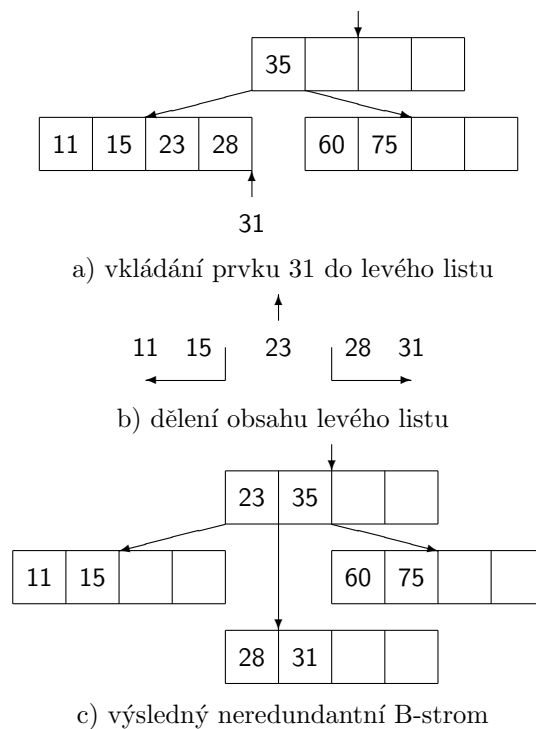
Obrázek 4: Doplnění prvků 11 a 15

existuje-li sousední list a má-li více prvků než minimální naplnění, jeden prvek si od něj vezmeme – podrobněji viz obr. ???. To ale nelze jen tak. Je třeba dodržet všechny podmínky, které neredundantní B-strom musí splňovat: klíč mezi dvěma odkazy do podstromů dělí množinu prvků v podstromech na prvky s menšími klíči než daný klíč a bna prvky s většími klíči, než je daný klíč. To bychom ale přímým převzetím prvku od souseda porušili. Proto je třeba provádět přesun prvku nepřímou: Prvek se od souseda přesouvá do rodičovského uzlu a dělicí prvek z rodičovského uzlu se přesouvá do uzlu podteklého. Tím se zajistí, že od dané chvíle struktura splňuje všechny podmínky kladené na neredundantní B-strom.

Pokud soused podteklého uzlu má jen minimum možných prvků, budeme oba uzly slévat (srv. obr. 7).

Je-li třeba vypustit prvek z nelistového uzlu neredundantního B-stromu, najde se v nejlevějším listu pravého podstromu nejmenší prvek nebo v nejpravějším listu levého podstromu největší prvek. Tímto nalezeným prvkem se nahradí prvek, který se měl vypustit, a nalezený prvek se ze svého původního místa užitím výše uvedeného postupu vypustí (obr. 10).

Spojíme prvky z obou uzlů s dělicí hodnotou z rodičovského uzlu. Tím ale může podteci rodičovský uzel. Podtečení na vyšších úrovních nederundantního B-stromu



Obrázek 5: Vložení prvku 31

se řeší obdobně jako podtečení listů. Je se v případě přesunu prvku od souseda přepojuje i jeden podstrom.

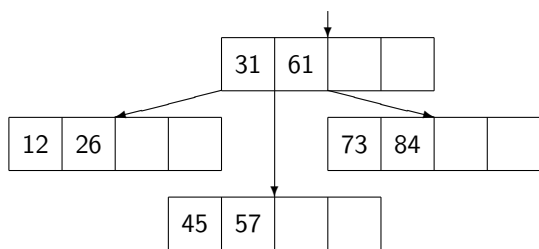
Postupným podtékáním uzlů může dojít až k situaci, kdy je výška stromu snížena – kdy zanikne jeho kořen. Přesněji: má-li kořen dva potomky a jeden z nich podteče, je tento sloučen s dělicím prvkem a obsahem druhého potomka. Tím je z kořene a jeho dvou potomků vytvořen nový kořen stromu – obr. 9.

Je-li třeba vypustit prvek z nelistového uzlu neredundantního B-stromu, najde se v nejlevějším listu pravého podstromu nejmenší prvek nebo v nejpravějším listu levého podstromu největší prvek. Tímto nalezeným prvkem se nahradí prvek, který se měl vypustit, a nalezený prvek se ze svého původního místa užitím výše uvedeného postupu vypustí (obr. 10).

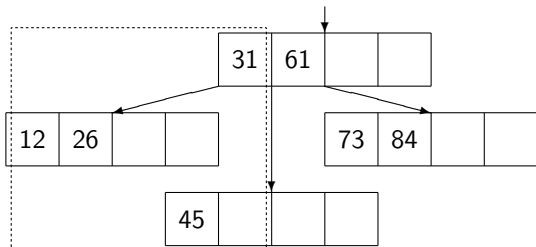
3 Redundantní B-stromy

Redundantní B-stromy ukládají všechna data v listech. Data uložená v nelistových uzlech jsou pouze pomocná. Nalezený klíč prvku je tedy platný pouze v listech. Najdeme-li jej pouze v nelistových uzlech, je to jako bychom jej vůbec nenašli.

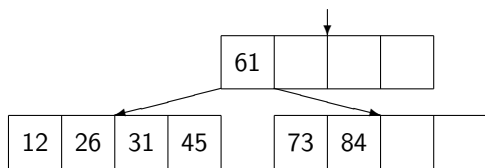
Tomu odpovídá i odlišné chování redundantních B-stromů při štěpení a dalších změnách na listové a nelistových úrovních. Také platí, že k vypouštění prvku vždy dochází pouze na listové úrovni – pomocné hodnoty zanikají při spojování uzlů, případně jsou měněny při přebírání prvku (a s ním případně spojeného podstromu) od souseda.



a) původní situace



b) podtečení uzlu po vypuštění prvku 57



c) výsledná podoba stromu

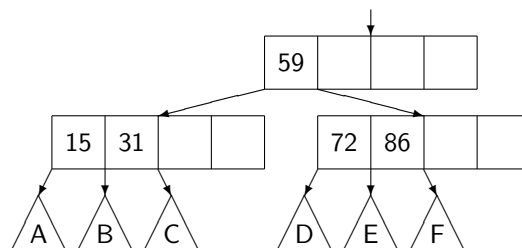
Obrázek 7: Sloučení uzlů neredundantního B-stromu

Obrázek 8: Slévání listů neredundantního B-stromu

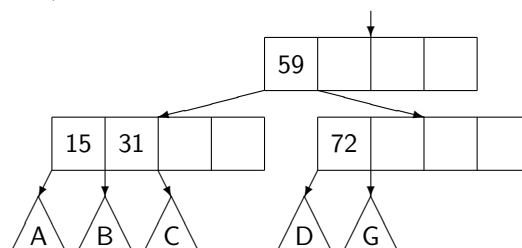
4 B⁺-stromy

Při využití redundantního B-stromu k indexaci dat, kdy listová úroveň ukazuje na data, zůstává v každém listovém uzlu jeden ukazatel nevyužitý. Těchto ukazatelů využijeme k tomu, aby si každý listový uzel mohl ukazovat na další list. Tím se umožní sekvenční průchod libovolné části listové úrovně stromu, aniž bychom potřebovali přístup k nelistovým uzlům – s výjimkou nalezení prvního prvku výběru.

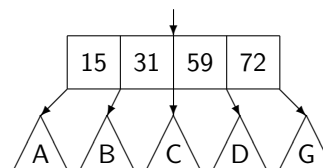
Prošíť listové úrovně B⁺-stromů umožňuje stejně rychlé vyhodnocování dotazů na jeden prvek jako klasický redundantní B-strom, ale navíc umožňuje rychlejší a snazší zpracování dotazů na sekence prvků. B⁺-stromy jsou vhodnou strukturou mj. pro indexy v databázích. Snadný přístup ke zvolenému prvku dle klíče a možnost efektivně sekvenčně pracovat s uloženými daty nabízí využití B⁺-stromu jako dynamické varianty *index-sekvenčního souboru*.



a) původní vrchní dvě úrovně B-stromu



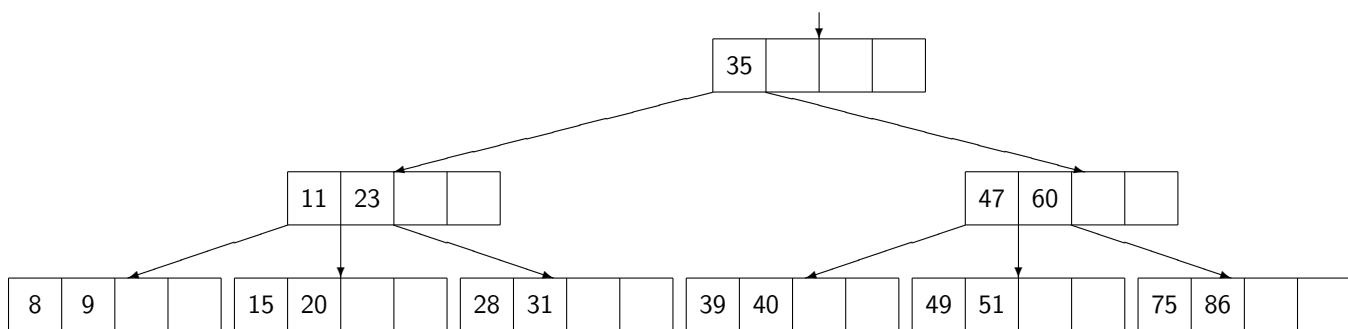
b) Po sloučení podstromů E a F (výsledek je G)



c) výsledná vrchní část stromu

Obrázek 9: Snížení úrovně neredundantního B-stromu

Obrázek 10: Vypuštění nelistového prvku z neredundantního B-stromu



Obrázek 6: B-strom se třemi úrovněmi