

5. BFák alkalmazása nagy adatbázisokban. Keresés, beszúrás, törlés megvalósítása

A **B-fák** olyan speciális fa adatstruktúrák, amelyek az adatokat rendezett formában tárolja, és megengedi a keresést, beszúrást és törlést legrosszabb esetben is logaritmikus időben (adatok mennyiségének növekedésével a beillesztés és törlés műveletigénye logaritmikusan nő.). Egy csomópont egy előre meghatározott tartományban változó mennyiségű leszármazottat tartalmazhat. Beillesztés és törlés esetén a csomópontok száma változik, és hogy a leszármazott csomópontok száma a meghatározott tartományon belül maradjon, lehetséges csomópont egyesítés és szétválasztás is. Egy B-fát szokás jellemezni azzal, hogy egy csomópontnak hány leszármazottja lehet. Egy m B-fa kielégíti a következő feltételeket:

- Minden csomópontnak legfeljebb m leszármazottja lehet
- Minden csomópontnak (kivéve a gyökeret és a leveleket) legalább $m/2$ leszármazottja van
- A gyökérnek legalább 2 leszármazottja van
- Minden levél azonos szinten van, és nem hordoznak információt
- Egy nem-levél csomópont, aminek k leszármazottja van, $k-1$ kulcsot tárol

A B-fák nagyon alkalmasak adatbázisokban és fílerendszerekben való felhasználásra. Mivel maximalizált a leszármazottak száma egy csomópontban, ezért a fa magassága csökken, kevesebbszer kell kiegyensúlyozni, és nő a hatékonysága. Általában a maximális leszármazottak számát a diszk blokkméretéhez szokták igazítani, vagy más, hasonló mérethez.

Műveletek B-fákon:

- Keresés: a keresés a bináris fákkal analóg módon működik, az ott bemutatott kétféle algoritmus itt is használható.
- Beszúrás: minden művelet a levél elemeken történik. Lépései:
 - A fán kereséssel meg kell találni az a csomópontot, ahova az új elem beilleszthetőHa a csomópontnak kevesebb leszármazottja van, mint a maximális megengedett, akkor oda beszúrható az új elem, de figyelembe kell venni, hogy rendezettek maradjanak az elemek
 - Más esetekben a levelet fel kell bontani két csomópontra:
 - Ki kell jelölni egy középelemet a régi és az új elemek közül
 - A mediánál kisebb elemek kerülnek az egyik új, a bal oldalon lévő levélre, a nagyobb elemek pedig a másik új csomópontra, a jobb oldalra. A medián mint elválasztó érték szerepel.
 - Az elválasztó érték hozzáadódik a szülő elemhez, ami abban a csomópontban okozhat szétválasztást, és így tovább, egészen a gyökér elemig (ha szükséges).
- Törlés: két féle megközelítés jellemző a törlésre:
 - kijelölni és törölni az elemet, majd újrastrukturálni a fát
 - végigmenni a fán egészen a kijelölt csomópontig, de még a csomópont elérése előtt újrastrukturálni a fát, hogy amikor törölve lesz az adott elem, utána már ne kelljen ezzel foglalkozni.

Két speciális eset lehet, amivel még törléskor foglalkozni kellhet:

- egy elem egy csomópontban elválasztó elemként szerepel: a bal oldali leszármazott legnagyobb és a jobboldali leszármazott legkisebb eleme közül valamelyiket ki kell jelölni, mint új elválasztót, és újra kell építeni az alatta lévő struktúrát
- egy elem törlésével a minimálisan szükségesnél kevesebb leszármazottja lesz egy elemnek: kiegyensúlyozás szükséges
- Kiegyensúlyozás: ha egy elem törlésével egy csomópontnak a kötelezően előírtnál kevesebb leszármazottja marad, akkor újra kell építeni a fát olyan módon, hogy mindenhol meglegyen a kötelező minimum. Bizonyos esetekben ez a művelet hiányt képez a szülő elemeknél, így az újraépítés eltarthat egészen a gyökérelemig is.