Algoritmuselmélet

2-3 fák

Katona Gyula Y.

Számítástudományi és Információelméleti Tanszék
Budapesti Műszaki és Gazdaságtudományi Egyetem

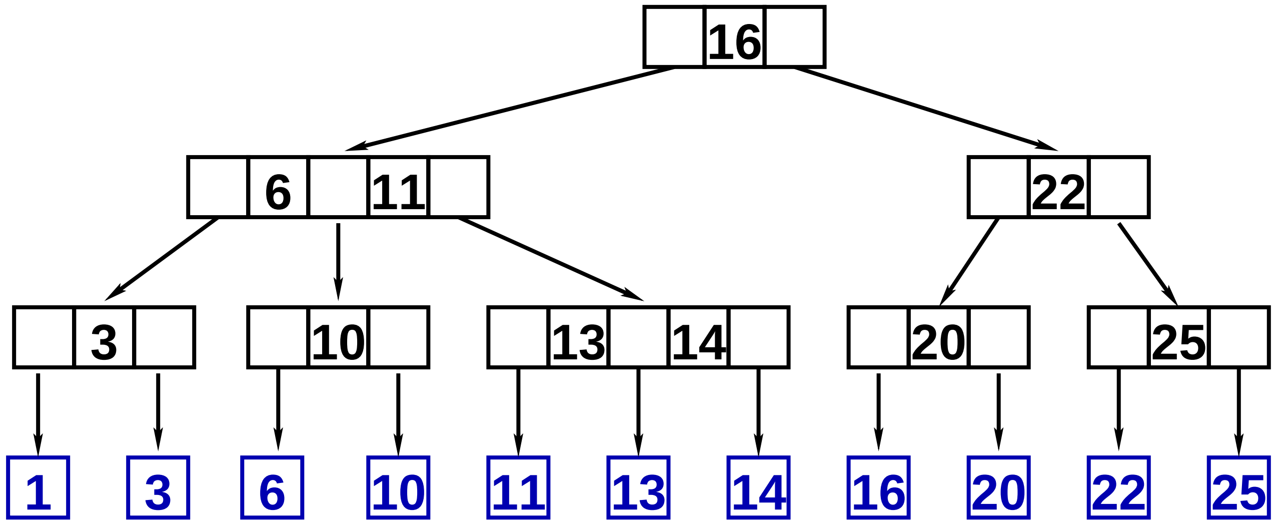
2-3-fák

Ez is fa, de a binárisnál bonyolultabb: egy nem-levél csúcsnak 2 vagy 3 fia lehet.

A **2-3-fa** egy (lefelé) irányított gyökeres fa, melyre:

- A rekordok a fa leveleiben helyezkednek el, a kulcs értéke szerint balról jobbra növekvő sorrendben. Egy levél egy rekordot tartalmaz.
- Minden belső (azaz nem levél) csúcsból 2 vagy 3 él megy lefelé; ennek megfelelően a belső csúcsok egy, illetve két $s \in U$ kulcsot tartalmaznak. A belső csúcsok szerkezete tehát kétféle lehet. Az egyik típus így ábrázolható: $\boxed{m_1 \mid s_1 \mid m_2 \mid s_2 \mid m_3}$
  Itt m_1, m_2, m_3 mutatók a csúcs részfáira, s_1, s_2 pedig U-beli kulcsok, melyekre $s_1 < s_2$. Az m_1 által mutatott részfa minden kulcsa kisebb, mint s_1, az m_2 részfájában s_1 a legkisebb kulcs, és minden kulcs kisebb, mint s_2. Végül m_3 részfájában s_2 a legkisebb kulcs. A másik típusú csúcsoknál az az utolsó két mező hiányzik: $\boxed{m_1 \mid s_1 \mid m_2}$
- A fa levelei a gyökértől egyforma távolságra vannak.

Példa 2-3-fára



2-3-fa tulajdonságai

Tétel

Ha a fának m szintje van, akkor a levelek száma legalább 2^{m-1}. *Megfordítva, ha* $|S| = n$ *(itt* $S \subseteq U$ *a fában tárolt kulcsok halmaza;* $|S|$ *megegyezik a tárolt rekordok számával), akkor* $m \leq \log_2 n + 1$.

Bizonyítás.

Minden belső csúcsnak legalább 2 fia van $\Longrightarrow$
az i-edik szinten legalább 2^{i-1} csúcs van ($1 \leq i \leq m$). $\Longrightarrow$
$2^{m-1} \leq n \Longrightarrow m - 1 \leq \log_2 n$. $\surd$ $\square$

2-3-fa tulajdonságai

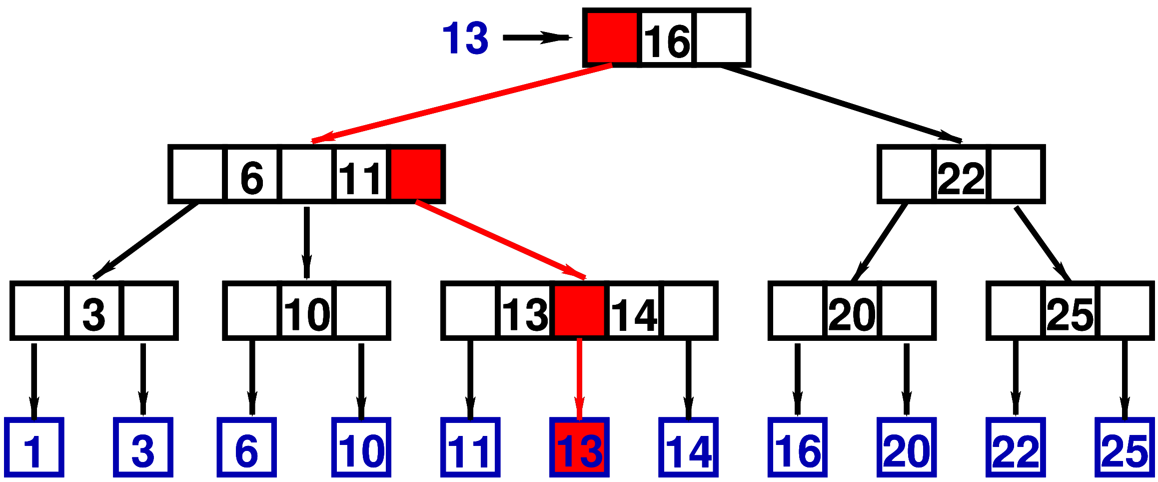
Tétel

Ha a fának m szintje van, akkor a levelek száma legfeljebb 3^{m-1}. Megfordítva, $m \geq \log_3 n + 1$.

Bizonyítás.

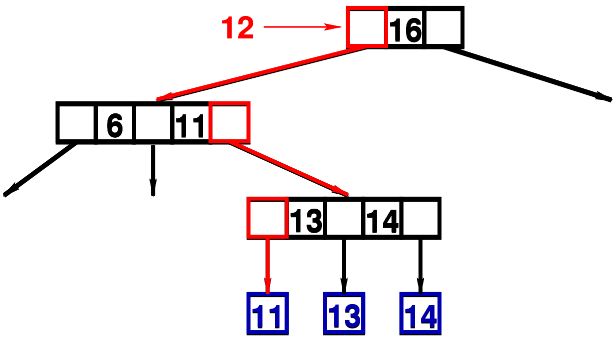
Minden belső csúcsnak legfeljebb 3 fia van $\Longrightarrow$
az i-edik szinten legfeljebb 3^{i-1} csúcs van ($1 \leq i \leq m$). $\Longrightarrow$
$n \leq 3^{m-1} \Longrightarrow m - 1 \geq \log_3 n$. $\surd$ $\square$

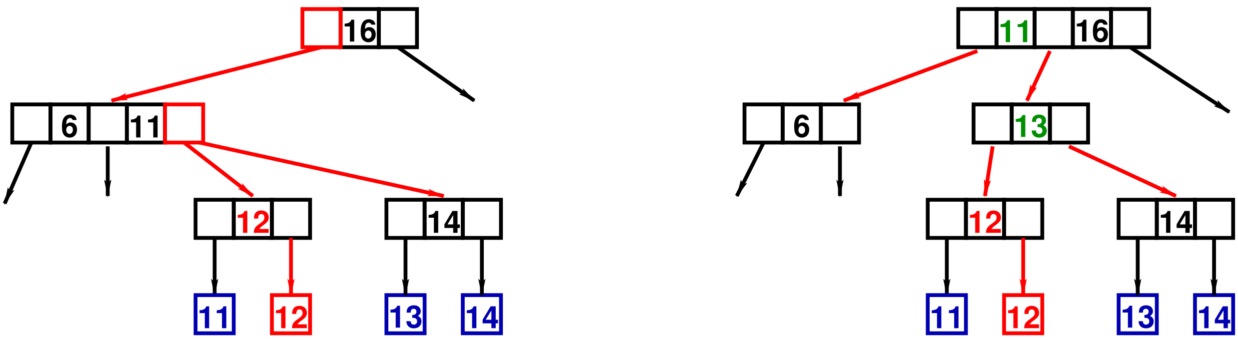
Keresés 2-3-fában



Hasonló, mint a bináris keresőfában.
Lépésszám: $O(m)$, ahol $\log_3(n) + 1 \leq m \leq \log_2(n) + 1$

BESZÚR 2-3-fába





Ha a gyökeret is „vágni” kell $\Longrightarrow$ új gyökér, nő a fa magassága.
Lépésszám: $O(m)$, minden szinten legfeljebb 1 „vágás”.

TÖRÖL 2-3-fából

Legyen x a legalsó belső csúcs a kereső út mentén.

- Ha x-nek három fia van $\Longrightarrow$ $\surd$
- Ha x-nek csak két fia van:
  - ha x (valamelyik) szomszédos testvérének 3 fia van $\Longrightarrow$ egyet átteszünk x alá;
  - ha x egyik szomszédos testvérének sincs három fia $\Longrightarrow$ „összevonunk” két kettes csúcsot.

Ez is „felgyűrűzhet”. $\Longrightarrow$

Lépésszám: $O(m)$

B-fák

R. Bayer, E. McCreight, 1972

A 2-3-fa általánosítása.

Nagy méretű adatbázisok, külső táron levő adatok feldolgozására használják. Több szabvány tartalmazza valamilyen változatát.

Probléma

Nem az összhasonlítás időigényes, hanem az adatok kiolvasása, de sokszor egy adat kiolvasásához amúgy is kiolvasunk több más adatot, egy lapot.

$\Longrightarrow$ A fa csúcsai legyenek lapok, a költség a lapelérések száma.

B-fa definíciója

Egy *m-edrendű B-fa*, röviden B_m-fa egy gyökeres, (lefelé) irányított fa, melyre érvényesek az alábbiaknak:

a) A gyökér foka legalább 2, kivéve esetleg, ha a fa legfeljebb kétszintes.

b) Minden más belső csúcsnak legalább $\lceil \frac{m}{2} \rceil$ fia van.

c) A levelek a gyökértől egyforma messze vannak.

d) Egy csúcsnak legfeljebb m fia lehet.

d) A tárolni kívánt rekordok itt is a fa leveleiben vannak; egy levélben a lapmérettől és a rekordhossztól függően több rekord is lehet, növekvően rendezett láncolt listában.

A belső csúcsok hasonlítanak a 2-3-fák belső csúcsaira. Egy belső csúcs így néz ki:

m_0	s_1	m_1	s_2	m_2	$\ldots$	s_i	m_i

A B-fa szintszáma

Tegyük fel, hogy egy B-fának n levele és k szintje van, és keressünk összefüggést e két paraméter között.
A kicsi fáktól eltekintve a gyökérnek legalább két fia van, a többi belső csúcsnak pedig legalább $\lceil \frac{m}{2} \rceil$.
$\Longrightarrow n \geq 2\lceil \frac{m}{2} \rceil^{k-2}$, $\Longrightarrow \log_{\lceil \frac{m}{2} \rceil} \frac{n}{2} + 2 \geq k$

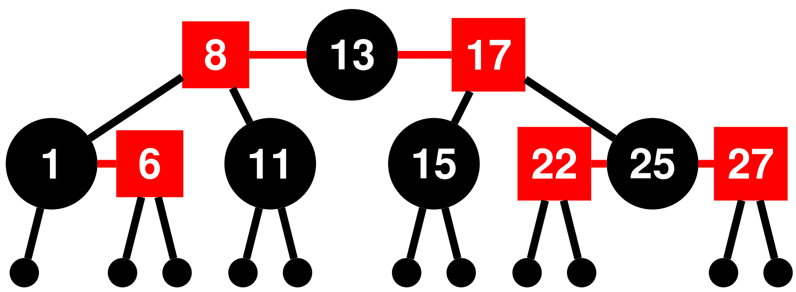
$$k \leq \frac{\log_2 n - 1}{\log_2 \lceil \frac{m}{2} \rceil} + 2.$$

Minden művelet lépésszáma: $\sim \frac{\log_2 n - 1}{\log_2 \lceil \frac{m}{2} \rceil} = \Theta\left(\frac{\log n}{\log m}\right)$, azaz a konstans szorzó kicsi, ha m nagy.
m viszont nem lehet túl nagy, hiszen a belső csúcsoknak egy lapon el kell férniük.

Például: Például, ha $m = 32$, $n = 2^{20}$ (itt n az alsó szint *lapjainak* száma), akkor $k \leq \frac{19}{4} + 2 < 7$. Egy rekord keresése tehát legfeljebb 6 lap elérését igényli.

A piros-fekete fa és a B-fa kapcsolata

A piros-fekete fa olyan B_4-fa, aminek a belső csúcsaiban tároljuk az elemeket.



A piros csúcsokat összevonjuk apjukkal, az így összevont csúcsoknak 2, 3 vagy 4 gyereke van.
Ezért a mélység csak a fekete csúcsokkal nő. $\Longrightarrow$ Mivel a fekete magasság állandó, minden levél azonos szinten lesz.

Java animáció: 2-3-fa.