

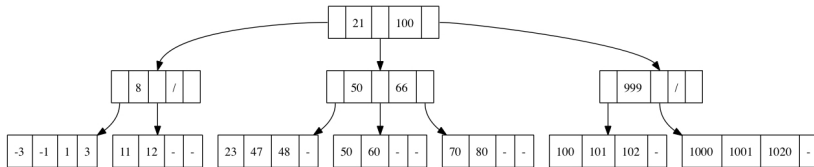
Indeksiranje podatkovnih baz

Osnove podatkovnih baz

2. letnik univerzitetnega študija na FRI

B⁺ indeks

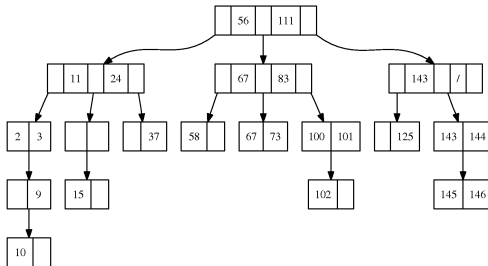
- ▶ B⁺ indeks je dinamično uravnoteženo drevo
- ▶ celotno drevo se prilagaja pri posodabljanju zapisov
- ▶ časovna zahtevnost osnovnih operacij $O(\log n) \ll O(n)$



B⁺ indeks za namišljeni atribut

ISAM indeks

- ▶ ISAM indeks je statično (ne)uravnoteženo drevo
- ▶ prelivne strani se prilagajajo pri posodabljanju zapisov
- ▶ časovna zahtevnost iskanja zapisov $\approx O(\log n) \ll O(n)$



ISAM indeks za namišljeni atribut

Bitni indeks

- ▶ bitni indeks je urejena binarna matrika
- ▶ celotna matrika se prilagaja pri posodabljanju zapisov
- ▶ časovna zahtevnost iskanja zapisov je *bitni* $O(n) < O(n)$

<i>id</i>	<i>name</i>	<i>year</i>	<i>color</i>	...
1	Ain	1966	blue	...
2	Ein	1950	red	...
3	Iin	1957	green	...
4	Oin	1950	blue	...
5	Uin	1950	red	...

Bitni indeks za atribut *year*

1950	1957	1966
0	0	1
1	0	0
0	1	0
1	0	0
1	0	0

Bitni indeks za atribut *color*

blue	green	red
1	0	0
0	0	1
0	1	0
1	0	0
0	0	1

1. naloga indeksiranja

d red drevesnega indeksa

b razvejanost drevesnega indeksa ($b = 2d + 1$)

h višina drevesnega indeksa (število nivojev brez listov)

a število zapisov v listih drevesnega indeksa

n število zapisov drevesnega indeksa

(a) Kdaj je smiselno uporabiti B^+ oziroma ISAM indeks?

(b) Najmanjše in največje število zapisov B^+ indeksa?

(c) Najmanjše in največje število zapisov ISAM indeksa?

└ 1. naloga indeksiranja

d red drevesnega indeksaa
 b razvejnost drevesnega indeksaa ($b := 2d + 1$)
 h višina drevesnega indeksaa (število nivojev brez listov)
 a število zapisov v listih drevesnega indeksaa
 n število zapisov drevesnega indeksaa

(a) Kdaj je smiselno uporabiti B^+ oziroma ISAM indeks?

(b) Najmanjše in največje število zapisov B^+ indeksa?

(c) Najmanjše in največje število zapisov ISAM indeksa?

Rešitve nalog:

(a) odvisno od primera uporabe

- B^+ (ISAM) indeks uporabimo, če so posodabljanja zapisov pogosta (redka/občasna)
- B^+ indeks uporabimo za spletno trgovino in ISAM indeks za trgovino, ki je ponoči zaprta

(b) $n \in \left[2^{\lceil \frac{a}{2} \rceil} (d + 1)^{h-1}, a(2d + 1)^h \right]$

(c) $n \in [1, \infty]$

2. naloga WikiLeaks

embassy (*id*, *name*)

privacy (*id*, *classification*)

cable (*id*, *date*, *header*, *content*, *#privacy_id*, *#embassy_id*)

reference (*#cable_id*, *#ref_cable_id*)

(a) Ustvarite tabelo *wiki* ter ji določite primarni ključ

```
CREATE TABLE wiki AS SELECT c.id, date, name, classification FROM cable c JOIN embassy e ON  
c.embassy_id = e.id JOIN privacy p ON c.privacy_id = p.id WHERE date > 0;
```

(b) Ocenite čas izvajanja in zahtevnosti naslednjih poizvedb

```
SELECT [SQL_NO_CACHE] * FROM wiki WHERE date >= '2009-01-01' AND date <= '2009-12-31';  
SELECT [SQL_NO_CACHE] * FROM wiki ORDER BY date;  
SELECT [SQL_NO_CACHE] MAX(date) FROM wiki;
```

Indeksiranje podatkovnih baz

2. naloga WikiLeaks

2. naloga WikiLeaks

```
embassy (id, name)
privacy (id, classification)
cable (id, date, header, content, #privacy_id, #embassy_id)
reference (#cable_id, #ref_cable_id)
```

(a) Ustvarite tabelo wiki ter ji določite primarni ključ

```
CREATE TABLE wiki (id, date, name, classification) PRIMARY KEY (id);
CREATE TABLE cable (id, date, header, content, #privacy_id, #embassy_id) PRIMARY KEY (id);
```

(b) Ocenite čas izvajanja in zahtevnosti naslednjih poizvedb

```
SELECT (id, date, content) FROM cable WHERE date >= '2009-01-01' AND date <= '2009-12-31';
SELECT (id, date, content) FROM cable WHERE date >= '2009-01-01' AND date <= '2009-12-31';
SELECT (id, date, content) FROM cable WHERE date >= '2009-01-01' AND date <= '2009-12-31';
```

Rešitve nalog:

- (a) `ALTER TABLE wiki ADD PRIMARY KEY (id);`
- (b) ≈ 100 milisekund in $O(n)$, $O(n \log n)$, $O(n)$

3. naloga WikiLeaks

embassy (*id*, *name*)

privacy (*id*, *classification*)

cable (*id*, *date*, *header*, *content*, *#privacy_id*, *#embassy_id*)

reference (*#cable_id*, *#ref_cable_id*)

(a) Pohitrite izvajanje zgornjih poizvedb z indeksiranjem

(b) Ustvarite indeksa za atributa *name* in *classification*

(c) Ali je pohitritev s pomočjo indeksiranja brezplačna?

3. naloga WikiLeaks

```
embassy (id name)
privacy (id classification)
cable (id date, header, content, #privacy_id, #embassy_id)
reference (#cable_id, #ref_cable_id)
```

(a) Pohitrite izvajanje zgornjih poizvedb z indeksiranjem

(b) Ustvarite indeksa za atributa name in classification

(c) Ali je pohitritev s pomočjo indeksiranja brezplačna?

Rešitve nalog:

- (a) `CREATE INDEX ind_date ON wiki(date);` ali `ALTER TABLE wiki ADD INDEX (date);`
- (b) v odjemalcu *phpMyAdmin* nad atributoma uporabi možnost *Index*
- (c) ne, časovna zahtevnost se izboljša na račun prostorske zahtevnosti

4. naloga WikiLeaks

embassy (*id*, *name*)

privacy (*id*, *classification*)

cable (*id*, *date*, *header*, *content*, *#privacy_id*, *#embassy_id*)

reference (*#cable_id*, *#ref_cable_id*)

(a) Pojasnite čas izvajanja naslednje poizvedbe

```
SELECT SQL_NO_CACHE * FROM wiki WHERE DATE(date) = '2009-12-10';
```

(b) Ustvarite skupen indeks za atributa *date* in *name*

(c) Poskusite ustvariti enoličen indeks za atribut *name*

4. naloga WikiLeaks

4. naloga WikiLeaks

```
embassy ((j, name))
privacy ((j, classification))
cable ((j, date, header, content, #privacy_id, #embassy_id))
reference ((#privacy_id, #embassy_id))
```

(a) Pojasnite čas izvajanja naslednje poizvedbe

```
SELECT * FROM CABLES WHERE DATE(date) < 2009-12-01;
```

(b) Ustvarite skupen indeks za atributa *date* in *name*

(c) Poskusite ustvariti enoličen indeks za atribut *name*

Rešitve nalog:

- (a) indeks je zgrajen preko vrednosti atributa *date* in ne *DATE(date)*
- (b) `CREATE INDEX ind_date_name ON wiki(date, name);`
- (c) `CREATE UNIQUE INDEX uni_ind_name ON wiki(name);`

5. naloga WikiLeaks

embassy (*id*, *name*)

privacy (*id*, *classification*)

cable (*id*, *date*, *header*, *content*, *#privacy_id*, *#embassy_id*)

reference (*#cable_id*, *#ref_cable_id*)

(a) Ustvarite tabelo *subwiki* ter ji določite primarni ključ

```
CREATE TABLE subwiki AS SELECT * FROM wiki WHERE date < '1975-12-31' ORDER BY date;
```

(b) Narišite bitna indeksa za atributa *name* in *classification*

(c) Kometirajte smiselnost bitnega indeksa za atribut *id*

5. naloga WikiLeaks

```
embassy ((j, name))
privacy ((j, classification))
cable ((j, date, header, content, #privacy_id, #embassy_id))
reference ((#embassy_id, #reference_id))
```

(a) Ustvarite tabelo subwiki ter ji določite primarni ključ

CREATE TABLE subwiki AS SELECT * FROM wiki WHERE id < 10701231 ORDER BY id;

(b) Narišite bitna indeksa za atributa name in classification

(c) Kometrajte smiselnost bitnega indeksa za atribut id

Rešitve nalog:

- (a) `ALTER TABLE subwiki ADD PRIMARY KEY (id);`
- (b) ...
- (c) bitni indeks ni smiseln, ker so vrednosti ključa enolične