

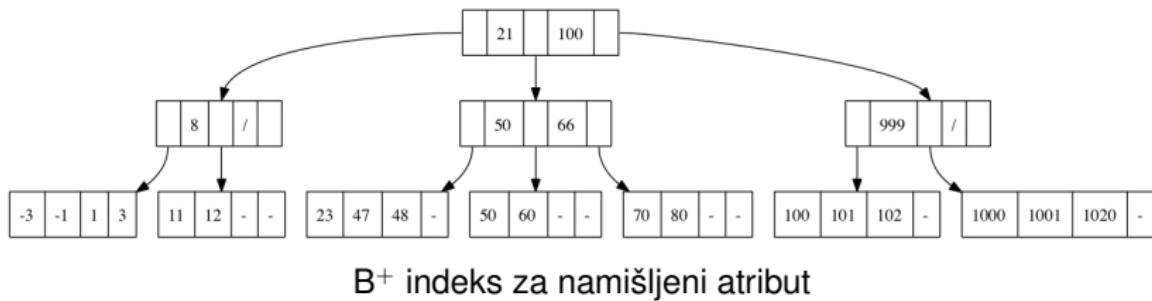
Indeksiranje podatkovnih baz

Osnove podatkovnih baz

2. letnik univerzitetnega študija na FRI

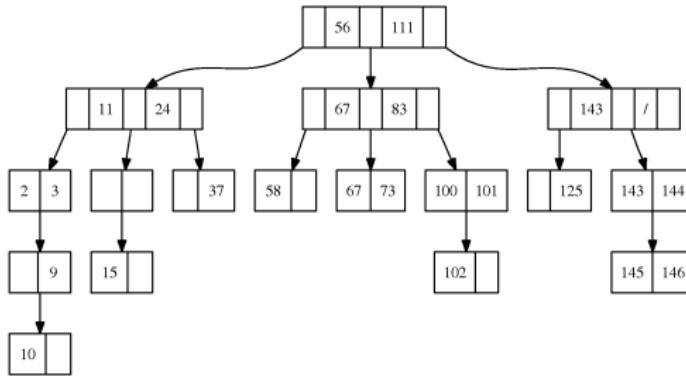
B⁺ indeks

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



ISAM indeks

- ▶ ISAM indeks je statično (ne)uravnoteženo drevo
 - ▶ prelivne strani se prilagajajo pri posodabljanu 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

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

d nad drevesnega indeksa
b razponost drevesnega indeksa ($b = 2d + 1$)
h veljna drevesnega indeksa (število nivojev brez listov)
a število zapris v listih drevesnega indeksa
n število zapris 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?

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

Rešitve nalog:

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

2. naloga WikiLeaks

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

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

```
CREATE TABLE wiki AS SELECT * FROM file_name classification PRIVACY table + JOIN embassy ON  
a embassy_id = id AND privacy_id = CN/privacy_id + JOIN reference table + R;
```

- (b) Ocenite čas izvajanja in zahtevnosti naslednjih poizvedb

```
SELECT file_id,(CN/CH) ,PRIVACY WHERE date >= '2009-01-01' AND date <= '2009-12-31'  
SELECT file_id,(CN/CH) ,PRIVACY ORDER BY date  
SELECT file_id,(CN/CH) ,PRIVACY WHERE file_id <
```

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?

Indeksiranje podatkovnih baz

└ 3. naloga WikiLeaks

3. naloga WikiLeaks

ambassy (# name)
privacy (# classification)
cable (# date, header, content, #privacy_id, #embassy_id)
reference (#wiki_id, #ref_cable_id)

- (a) Pohtrite izvajanje zgornjih poizvedb z indeksiranjem
- (b) Ustvarite indeksa za atributa name in classification
- (c) Ali je pohtritev 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.NOCACHE * 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*

Indeksiranje podatkovnih baz

4. naloga WikiLeaks

└ 4. naloga WikiLeaks

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);*

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

(a) Pojasnite čas izvajanja naslednje poizvedbe
SELECT SOLICITATION_ID FROM wiki WHERE DATE(date) = '2008-12-12'

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

(c) Poskusite ustvariti enoličen indeks za atribut *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

Rešitve nalog:

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

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

(a) Ustvarite tabelo subwiki ter ji določite primarni ključ

```
CREATE TABLE subwiki AS SELECT * FROM wiki WHERE date < '1979-12-31' ORDER BY date;
```

(b) Narišite bitna indeksa za atributa name in classification

(c) Komentirajte smiselnost bitnega indeksa za atribut id