



UNIVERSITETET
I OSLO

Bitmap-indekser

Bitmap-indekser

- Utgangspunkt: Alle recorder er tilordnet et uforanderlig, entydig tall
 - Nummering fra 1 til n
 - Nummeret kan betraktes som en record-ID og kan ikke gjenbrukes
- Velg ut feltet F som det skal lages en indeks på
 - For hver benyttet verdi v for F i en av recordene, opprett en bitvektor b_v med lengde n
 - Hvis record nr. i har $F=v$, la $b_v[i]=1$
 - Hvis record nr. i har $F \neq v$, la $b_v[i]=0$

Eksempel

record-
nummer

	fil	
	F	G
1	30	foo
2	30	bar
3	40	baz
4	50	fou
5	40	bar
6	30	baz

bitvektorer for F

30:	1	1	0	0	0	1
40:	0	0	1	0	1	0
50:	0	0	0	1	0	0

Karakteristika

- Plassbehov:
 - Totalt antall bits er $\#records * \#verdier$
 - I verste fall trengs n^2 bits (men da har hver bitvektor bare én 1-bit)
 - Bitvektorene kan komprimeres; det er aldri mer enn n 1-bits totalt i bitvektorene
- Effektiv for
 - partial match queries (= angi verdier for noen felter, finn alle som har gitte verdier)
 - Beregn bitvis AND på tvers av bitmap-indeksene for de aktuelle attributtene
 - range queries (= angi intervaller for noen felter, finn alle som har verdier innen intervallene)
 - Beregn bitvis OR innen intervallene og bitvis AND mellom intervallene

Kompresjon

- En *run* er en sekvens av nuller etterfulgt av en ener
- Representer hver run ved en kodebit:
 - Gitt en run av i nuller (og én ener)
 - Finn hvor mange sifre j som trengs til å representere i binært
 - Kod run'en ved $j-1$ enere, fulgt av null, fulgt av kodingen av i minus i 's innledende ener (bare hvis $j>1$)
- Konkatener alle kodede runs i en bitvektor. Dropp avsluttende nuller.
- Kan maksimalt komprimere til $2\log_2 n$ bits pr. bitvektor

Eksempel

0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

001 kodes til ($i=2=\underline{10}_2$, $j=2$) 100

01 kodes til ($i=1=\underline{1}_2$, $j=1$) 01

1 kodes til ($i=0=\underline{0}_2$, $j=1$) 00

00000000001 kodes til ($i=15=\underline{1111}_2$, $j=4$) 1110111

00 droppes

Totalt: 10001001110111