

Índices invertidos

ALGORITMOS E ESTRUTURAS DE DADOS III

Prof. Marcos André S. Kutova

Índices invertidos

- Um índice invertido é um índice em que uma parte do conteúdo de um registro (como uma palavra de um campo) é usada na localização do próprio registro.
- Essa é uma solução para permitir, entre outras, a busca de texto em arquivos, como fazem as máquinas de busca na Web.

Índices invertidos

Cód.	Título	...
1	Implementação de sistemas de bancos de dados	...
2	Sistemas de bancos de dados	...
3	Estruturas de dados e seus algoritmos	...
4	Dominando algoritmos	...
5	Estruturas de dados em Java	...
6	Core Java	...
7	Biblioteca do programador Java	...

The diagram illustrates inverted indices for the word "dados". Four red curved arrows originate from the word "dados" and point to the occurrences of the word in the table rows. The arrows point to the following rows:

- Row 1: Implementação de sistemas de bancos de dados
- Row 2: Sistemas de bancos de dados
- Row 3: Estruturas de dados e seus algoritmos
- Row 5: Estruturas de dados em Java

The word "dados" is highlighted in red in the original image.

Busca sequencial ?

- A busca sequencial (testando cada registro) é lenta demais para um grande volume de dados
 - Google: pesquisa em bilhões de páginas em uma fração de segundo

Índice invertido

- Todos os "termos" são identificados e, para cada um deles, criamos uma lista dos registros em que aparecem
- Índice invertido = listas invertidas

Índice invertido

Cód.	Título	...
1	Implementação de sistemas de bancos de dados	...
2	Sistemas de bancos de dados	...
3	Estruturas de dados e seus algoritmos	...
4	Dominando algoritmos	...
5	Estruturas de dados em Java	...
6	Core Java	...
7	Biblioteca do programador Java	...

Termos
implementação
de
sistemas
bancos
dados
estruturas
e
seus
algoritmos
dominando
em
java
core
biblioteca
do
programador

Retirar termos não
significativos e
ordenar conjunto

Índice invertido

Cód.	Título	...
1	Implementação de sistemas de bancos de dados	...
2	Sistemas de bancos de dados	...
3	Estruturas de dados e seus algoritmos	...
4	Dominando algoritmos	...
5	Estruturas de dados em Java	...
6	Core Java	...
7	Biblioteca do programador Java	...

Termos	Registros
algoritmos	3 4
bancos	1 2
biblioteca	7
core	6
dados	1 2 3 5
dominando	4
estruturas	3 5
implementação	1
java	5 6 7
programador	7
seus	3
sistemas	1 2

Exemplo de consulta

Termos	Registros
algoritmos	3 4
bancos	1 2
biblioteca	7
core	6
dados	1 2 3 5
dominando	4
estruturas	3 5
implementação	1
java	5 6 7
programador	7
seus	3
sistemas	1 2

- Consulta: "estruturas ~~de~~ dados"
- Primeiro termo: "estruturas"
Conjunto: { 3, 5 }
- Repetir:
 - Próximo termo: "dados"
 - Conjunto teste: { 1, 2, 3, 5 }
 - Conjunto atual: { 3, 5 }
 - Conjunto resposta: { 3, 5 }
(interseção)

Vantagens

- Os índices invertidos (ou listas invertidas) podem ser construídos para qualquer conjunto de informações dos registros.
- Esses índices são estruturas adequadas para consultas combinadas (vários campos).