

Algoritmos e Estruturas de Dados III

Aula 6 - Multilistas e Índices Invertidos

Prof. Felipe Lara



PUC Minas

Índices Secundários

- Estruturas de dados tipo arquivo **não** são usadas estritamente para disponibilizar consultas a partir de chaves primárias, que identificam unicamente os registros

Índices Secundários

- Estruturas de dados tipo arquivo **não** são usadas estritamente para disponibilizar consultas a partir de chaves primárias, que identificam unicamente os registros
- **Chaves Secundárias:** são chaves cujos valores identificam mais de um registro
 - diferente das chaves primárias, a partir das quais obtém-se um único registro
- Consultas cujas respostas envolvem **mais de um registro** são formuladas a partir de possíveis chaves secundárias

Índices Secundários

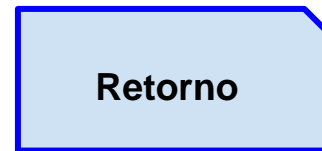
- Vários problemas apresentam chaves com valores idênticos (elas não distinguem unicamente os registros)
- As buscas resultam em subconjuntos de registros

Índices Secundários

- Vários problemas apresentam chaves com valores idênticos (elas não distinguem unicamente os registros)
- As buscas resultam em subconjuntos de registros

Exemplo: busca por idade = 18

Posição	Nome	Idade	Cidade
01	João	19	BH
02	Ana	18	RJ
03	Pedro	19	SP
04	Maria	20	BH
05	Carlos	18	SP

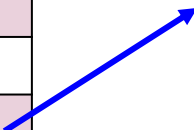


Índices Secundários

- Vários problemas apresentam chaves com valores idênticos (elas não distinguem unicamente os registros)
- As buscas resultam em subconjuntos de registros

Exemplo: busca por idade = 18 e cidade = SP

Posição	Nome	Idade	Cidade
01	João	19	BH
02	Ana	18	RJ
03	Pedro	19	SP
04	Maria	20	BH
05	Carlos	18	SP



Retorno

Índices Secundários

Um arquivo em que são feitas buscas a partir de chaves secundárias deve ser organizado para minimizar o esforço de pesquisa.

Índices Secundários

Um arquivo em que são feitas buscas a partir de chaves secundárias deve ser organizado para minimizar o esforço de pesquisa. Possibilidades:

Arquivos/Índices multilista: mantém-se um link para cada chave secundária, sendo que os registros que apresentam o mesmo valor para essa chave são encadeados.

Arquivos/Índices invertidos: mantém-se arquivos adicionais que indexam os registros que apresentam mesmo valor de chave secundária

Multilistas



PUC Minas

Multilistas

- Arquivos representam links para cada chave secundária, usando ponteiros para encadear os registros que apresentam o mesmo valor de chave secundária.
- Para cada chave secundária é elaborado um índice correspondente e todos os índices são mantidos numa área denominada diretório

Multilistas

Arquivos Multilistas são compostos por:

- diretório contendo 1 ou mais índices
- área de registro de dados

Multilistas

Posição	Nome	Idade	Link I	Cidade
01	João	19	03	BH
02	Ana	18	05	RJ
03	Pedro	19	-1	SP
04	Maria	20	-1	BH
05	Carlos	18	-1	SP

Idade	Quantidade	Pos. Inicial
18	02	02
19	02	01
20	01	04

Multilistas

Posição	Nome	Idade	Link I	Cidade
01	João	19	03	BH
02	Ana	18	05	RJ
03	Pedro	19	-1	SP
04	Maria	20	-1	BH
05	Carlos	18	-1	SP

Idade	Quantidade	Pos. Inicial
18	02	02
19	02	01
20	01	04

Multilistas

Posição	Nome	Idade	Link I	Cidade
01	João	19	03	BH
02	Ana	18	05	RJ
03	Pedro	19	-1	SP
04	Maria	20	-1	BH
05	Carlos	18	-1	SP

Idade	Quantidade	Pos. Inicial
18	02	02
19	02	01
20	01	04

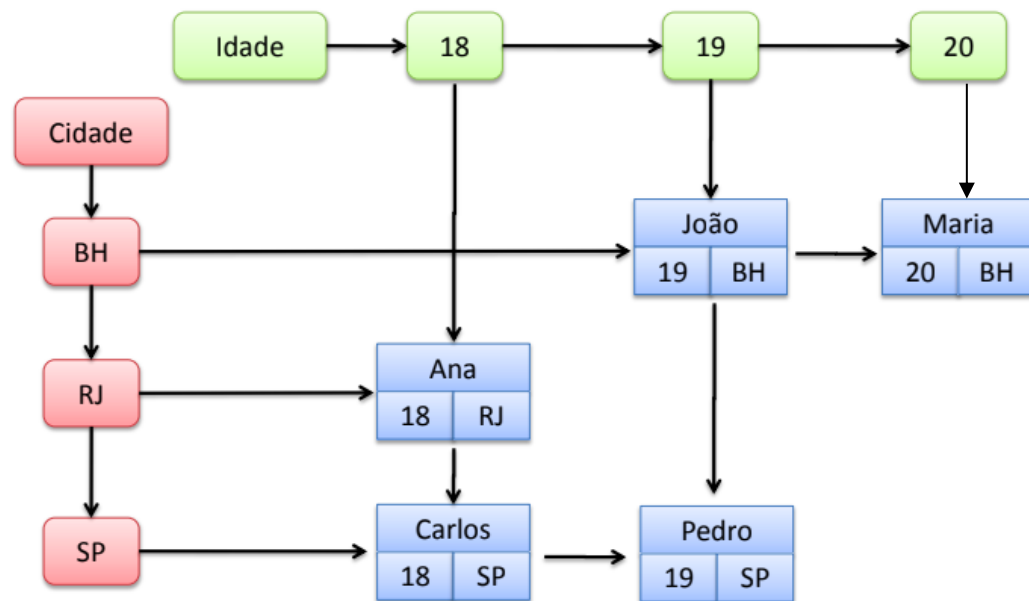
Multilistas

Posição	Nome	Idade	Link I.	Cidade	Link C.
01	João	19	03	BH	04
02	Ana	18	05	RJ	-1
03	Pedro	19	-1	SP	05
04	Maria	20	-1	BH	-1
05	Carlos	18	-1	SP	-1

Idade	Quantidade	Pos. Inicial
18	02	02
19	02	01
20	01	04

Cidade	Quantidade	Pos. Inicial
BH	02	01
RJ	01	02
SP	02	03

Multilistas



Posição

01

Nome

João

Idade

19

Cidade

BH

02

Ana

18

RJ

03

Pedro

19

SP

04

Maria

20

BH

05

Carlos

18

SP

Posição	Nome	Idade	Cidade
01	João	19	BH
02	Ana	18	RJ
03	Pedro	19	SP
04	Maria	20	BH
05	Carlos	18	SP

Índices invertidos



Índices invertidos

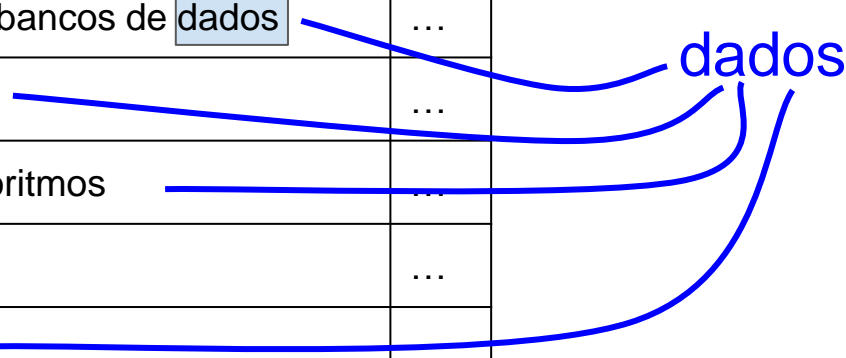
- Um índice invertido é um índice em que uma parte do conteúdo de um registro (como uma palavra de um campo) é usada na localização do próprio registro.
- Essa é uma solução para permitir, entre outras, a busca de textos em arquivos, como fazem as máquinas de busca na Web.

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	...

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	...



dados

Busca sequencial ?

- 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

Aproximadamente 14.600.000.000 resultados (0,82 segundos)

https://pt.wikipedia.org/wiki/House,_M.D.

House, M.D. – Wikipédia, a enciclopédia livre

Gregory House, interpretado pelo ator inglês Hugh Laurie. A série passa-se no fictício hospital universitário Princeton-Plainsboro Teaching Hospital, em ...

Temporadas: 8

Episódios: 177 (Lista de episódios)

Elenco: Hugh Laurie; Lisa Edelstein; Omar Ep... Tema de abertura: "Teardrop"

Gregory House · Lista de episódios de House... · Lisa Edelstein · Kal Penn

📰 Principais notícias

Notícias sobre A Casa do Dragão



TecMundo

House of the Dragon: fãs encontram erro de CGI no 3º episódio; veja!

5 horas atrás

GZH

Com caprichos de Rhaenyra e batalha sangrenta, saiba como foi o terceiro episódi...

10 horas atrás

IG Queer

'House of the Dragon' apresenta primeiro personagem gay

10 horas atrás



House, M.D.

2004 · Drama · 8 temporadas



Assistir agora
Assinatura



Assisti



Quero assistir

▼ Todas as opções para assistir

89% gostaram desse programa de TV

Usuários do Google



No fictício Hospital-Escola Plainsboro de Princeton,

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

Índice Invertido

Retirar termos
não significativos
e ordenar
conjunto

Termos

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

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	...

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
sistemas
bancos
dados
estruturas
seus
algoritmos
dominando
java
core
biblioteca
programador

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

algoritmos
bancos
biblioteca
core
dados
dominando
estruturas
implementação
java
programador
seus
sistemas

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

Índice Invertido - Consultas

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}**

Índice Invertido - Consultas

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**}

Índice Invertido - Consultas

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}

Unindo os dois:

Conjunto atual: {3,5} \cap {1,2,3,5}

➤ Resposta: **{3,5}**

Índice Invertido - Consultas

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	...

Resposta: {3,5}

Índice Invertido - Outro Exemplo

Posição	Nome	Idade	Cidade
01	João	19	BH
02	Ana	18	RJ
03	Pedro	19	SP
04	Maria	20	BH
05	Carlos	18	SP

Inversão Por Idade

Idade	Referências
18	02 05
19	01 03
20	04

Inversão Por Cidade

Cidade	Referências
BH	01 04
RJ	02
SP	03 05

Índice Invertido - 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).

Implementação relativamente fácil

É a estrutura de dados mais popular usada em sistemas de recuperação de documentos, usada em larga escala, por exemplo, em mecanismos de busca