

Algoritmos e Estruturas de Dados III

Aula 6.3 – Multilistas e Índices Invertidos

Prof. Hayala Curto
2022



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
- Consultas cujas respostas envolvem mais de um registro são formuladas a partir de possíveis chaves secundárias
- Chaves Secundárias são chaves cujos valores identificam mais de um registro, diferente das chaves primárias, a partir das quais obtém-se um único registro

Índices Secundários

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

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

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 C, através dos quais, registros que apresentam mesmo valor de C 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 representem 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

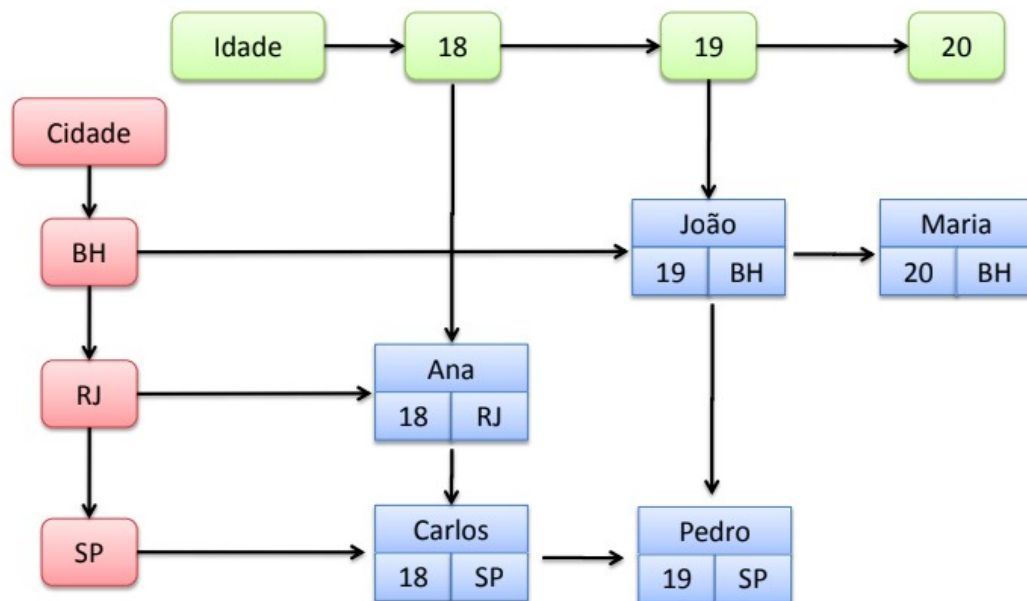
Multilistas

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

Índice Invertido



Índice Invertido

- 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.

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	...

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	...

Dados

Índice Invertido

- 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

The screenshot shows a Google search interface with the query 'house'. The search results are displayed in a grid-like format. The top result is for 'House, M.D.' from Wikipedia, with a blue arrow pointing to the URL. Below it are 'Principais notícias' (Main news) and a section for 'House of the Dragon' featuring a character from the show. On the right side, there is a dedicated section for 'House, M.D.' with a large image of Hugh Laurie, a description of the show, and options to 'Assistir agora' (Watch now), 'Assisti' (Watched), or 'Quero assistir' (I want to watch). The bottom of the right section shows a rating of 89% and a list of users who watched the program.

Google

house

Todas Imagens Vídeos Maps Notícias Mais Ferramentas

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

House, M.D.

2004 · Drama · 8 temporadas

Assistir agora Assisti Quero assistir

Todas as opções para assistir

89% gostaram desse programa de TV

Usuários do Google

No fictício Hospital-Escola Plainsboro de Princeton,

Patente
Google

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

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
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
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: “estrutura **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)

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

Índice Invertido - Desvantagens

- Grande sobrecarga de armazenamento e altos custos de manutenção na atualização, exclusão e inserção.

Índice Invertido - TP1

Orientações sobre a criação da lista invertida.

- Deve-se criar um arquivo contendo a lista invertida para o nome da pessoa.
- Deve-se criar um arquivo contendo a lista invertida para a cidade da pessoa.
- O sistema deverá realizar alterações nas listas invertidas sempre que novos registros forem inseridos, alterados ou deletados no arquivo de dados.
- O sistema deve ser capaz de receber uma busca por nome e/ou cidade.
- A partir do nome e/ou cidade digitados, o sistema deve ser capaz de retornar em quais ids de registros a pesquisa foi capaz de encontrar os termos informados.