

Professor: Rosane Minghim
Assistente PAE: Renato Rodrigues
Monitor: Fernando Bittencourt

Arquivos - Índices e Compressão

Contextualização

Uma empresa possui um sistema de vendas on-line. Nesse sistema diversos produtos estão cadastrados em uma base de dados, utilizada para consultas, construção de relatórios e movimentação de estoque. Ao ser desenvolvido, esse sistema não estava preparado para suportar o crescente volume de transações da empresa, e o administrador sugeriu algumas mudanças na próxima versão.

As consultas à base de dados devem ser aceleradas para que os clientes não fiquem esperando muito tempo para obter as informações sobre os produtos. Outra sugestão do administrador é utilizar algum método de compressão de dados em algumas tabelas do sistema, que passaram a ocupar muito espaço no servidor da empresa.

Para acelerar a consulta à base de dados, a equipe de desenvolvimento sugeriu adotar novos índices nas bases de dados. Esses índices devem indicar o local exato dos registros nos arquivos, sem a necessidade de percorrer toda a base para encontrar os itens pesquisados pelos clientes. Para otimizar a ocupação de espaço nas tabelas, a equipe de desenvolvimento decidiu aplicar o método de Huffman para comprimir os campos que ocupam mais espaço na base de dados.

Especificação

Para verificar a viabilidade dessa solução, a equipe de desenvolvimento decidiu realizar testes sobre a tabela de descrição de produtos do sistema. Essa tabela possui apenas dois campos: nome do produto e descrição do produto. Ambos são campos string de tamanho fixo. O nome de cada produto é único, e pode conter até 50 caracteres. A descrição do produto pode conter até 100 caracteres.

Dado um conjunto de registros da tabela de descrição de produtos, desenvolva um programa para criar um índice para cada registro. O índice é um número que identifica qual é a posição do registro. Desenvolva também uma função para aplicar o algoritmo de Huffman para comprimir o campo de descrição do produto.

Dados de entrada e saída

O SQTPM deve informar o conjunto de registros, e o programa deve ser capaz de informar qual é o índice (número identificador do registro na sequência) para um dado registro identificado pelo nome. O SQTPM também deverá consultar qual foi o código Huffman gerado para um dado caractere presente no campo de descrição do produto.

Os dados de entrada do programa, informados pelo SQTPM:

- O conjunto de registros da tabela de descrição de produtos (string).
- O nome identificador de um registro (string).
- Um caractere presente na descrição do registro (string de um caractere).

Os dados de saída do programa, usados pelo SQTPM:

- O índice do registro identificado no segundo dado de entrada (inteiro).
- O código Huffman do caractere informado no terceiro dado de entrada (string).

O primeiro dado de entrada é o conjunto de registros. O conjunto de registros da tabela é informado pelo SQTPM como um único string. Os registros são separados pelo caractere '#', e os campos de cada registro são separados pelo caractere ';'. O programa deve ler esse conjunto de caracteres e separar os registros e os campos de cada registro. O programa deve ser capaz de criar um número sequencial para indexar cada registro na ordem em que aparecem.

O segundo dado de entrada é o nome identificador de um registro. Esse nome é único para cada registro. A partir dessa informação, o programa deve ser capaz de encontrar o registro com o mesmo nome informado no conjunto. O terceiro dado de entrada é um string de um único caractere. Esse caractere está presente no campo de descrição do registro.

A partir dessas informações, o programa deve informar, para o registro encontrado, qual é o seu índice e código Huffman para o caractere informado.

Exemplo

Dados de entrada:

- registro1;descricao do registro1#registro2;descricao do registro2#registro3;descricao do registro3
- registro3
- e

Saída esperada:

- 3
- 101

Para produzir a saída, utilize o comando: `printf("%d %s\n", indiceRegistro, codigoCaractere);`

Prazo máximo de entrega: 11/06/2013