

ICMC-USP  
SCC0603 – Algoritmos e Estruturas de Dados II  
Prof. Rosane Minghim – 1º Semestre 2012

**Trabalho 4 – Árvore B**

Esse trabalho consiste na criação e atualização de um índice de chaves como Árvore B. Assim como no trabalho anterior (3), esse índice pode ser utilizado para facilitar a manipulação de registros de um arquivo muito grande para ser mantido completamente na memória primária. Novamente implementaremos uma versão simples de um conceito muito utilizado na prática em sistemas de gerenciamento de bancos de dados.

A entrada é similar à do trabalho 3 e consistirá em um conjunto de referências bibliográficas (o nosso banco de dados) seguido de um conjunto de operações a serem realizadas nesses dados: adição ou remoção de registros. Nesse caso, no entanto, você deverá imprimir como saída a Árvore B resultante das operações.

## 1 Detalhes da entrada

Abaixo segue um exemplo de entrada e a explicação de cada parte:

```
3
2
Leung, Joseph Y.-T./A new algorithm for scheduling periodic, real-time
  tasks/Algorithmica/4/2/1989/209-219/0178-4617
Jansen, Klaus/Scheduling Malleable Parallel Tasks An Asymptotic Fully
  Polynomial Time Approximation Scheme/Algorithmica/39/1/2004/59-81/
  0178-4617
2
ADD/Anderson, Richard J. and Miller, Gary L./Deterministic Parallel List
  Ranking/Algorithmica/6/6/1991/859-868/0178-4617
REMOVE/0178-4617/39/1/59-81
```

A primeira linha fornece ordem  $D$  da árvore. A segunda linha fornece o número  $N$  de referências que formam o banco de dados. As  $N$  linhas seguintes são as referências; cada referência possui vários campos, que estão no formato:

```
author/title/journal/volume/number/year/pages/issn
```

Observe que não há espaços em branco ou quaisquer outros símbolos separando cada campo, apenas o `"/`. Além disso, não há quebras nas linhas (aqui no enunciado as linhas

estão quebradas para que caibam na página). O ISSN é um número que identifica de forma única um *journal* no mundo.

A primeira linha após as  $N$  referências fornece o número  $M$  de operações a serem realizadas no banco de dados. As  $M$  linhas seguintes são as operações, que podem ser de dois tipos:

1. **Adicionar registro:** A operação de adição de registro é identificada por `ADD`, seguida de uma referência completa no mesmo padrão mostrado acima.
2. **Remover registro:** A operação de remoção de registro é identificada por `REMOVE`, seguida de uma sequência "ISSN/volume/number/pages" que identifica (de forma única) o registro a ser removido.

## 2 Detalhes da saída esperada

Você deverá imprimir o índice (Árvore B) de volta na saída padrão após a realização de todas as operações indicadas. Cada nó (página) da árvore deve ser impresso em uma linha, precedido pelo seu nível. Consideraremos a raiz como sendo nível 0, seus filhos como nível 1 e assim por diante. A ordem deve seguir uma busca por largura, ou seja, todos os nós de nível 1, seguidos de todos os nós de nível 2, etc. Além disso, entre os nós de um mesmo nível, deve ser seguida a ordem crescente das chaves. Por fim, cada linha (nó) é composta pelas chaves em ordem crescente e separadas por vírgulas.

Um exemplo de saída de uma árvore de ordem 3 (não equivalente ao exemplo de entrada) seria:

```
0 0178-4617/48/1/37-66
1 0178-4617/21/1/119-136, 0178-4617/42/1/1-2
1 0178-4617/55/4/666-702, 0178-4617/62/3/982-1005
```

## 3 Implementação

Os recursos disponíveis (memória e processamento) para o trabalho serão extremamente limitados; ou seja, não será possível carregar todos os dados na memória para realizar as operações. Você deverá implementar o índice na memória carregando as referências/operações uma de cada vez.

## 4 Observações

- A entrada foi formatada especificamente para facilitar a leitura com a função `strtok`. Exemplo:

```
fgets(line, MAX_LINE_CHARS, stdin);
author = strtok(line, "/");
title = strtok(NULL, "/");
...
```