

# MAC 5710 – Estrutura de Dados e sua Manipulação

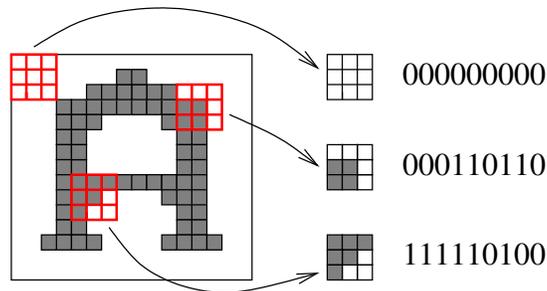
Primeiro semestre de 2007

Exercício-Programa 3 (EP3) — Data de entrega: 21/06/2007

**Objetivos:** Implementar e avaliar o desempenho de árvores binárias de busca (com e sem balanceamento) para dados obtidos de imagens binárias.

**Obtenção dos dados:** Os dados a serem inseridos nas árvores deverão ser obtidos a partir de uma imagem binária. Seja  $I$  uma imagem binária de dimensões  $m \times n$  ( $m$  linhas e  $n$  colunas) e  $W$  uma “janela” de dimensões  $p \times q$  (retângulo com  $p$  linhas e  $q$  colunas). Considere ainda  $p < m$  e  $q < n$ . Ao se posicionar  $W$  sobre algum ponto  $(i, j)$  de  $I$ , se  $W$  estiver inteiramente sobre  $I$ , podemos observar uma subimagem  $I'$  de  $I$  restrita a  $W$ , de dimensões  $p \times q$ . A representação vetorial dessa subimagem, obtida concatenando-se uma linha após a outra em uma única linha, será denominado um padrão.

**Exemplo:** Na figura a seguir, ao lado esquerdo encontra-se uma imagem binária e, ao lado direito, três padrões de tamanho 9 obtidos de três diferentes posições na imagem.



Uma vez que todos os valores observados sob a janela são ou zero ou um, um padrão pode ser encarado como um número binário de  $d = p \cdot q$  bits. Logo, pode-se estabelecer uma relação de ordem total entre os possíveis  $2^d$  padrões.

**Árvores a serem implementadas:** os dados obtidos deverão ser inseridos em uma árvore binária de busca sem balanceamento e em uma AVL ou RB-Tree. Você pode escolher uma entre a AVL e a RB-Tree. Além dos campos usuais necessários para manipular a árvore, os nós da árvore deverão possuir um campo **freq** que corresponde à frequência de ocorrência do respectivo padrão armazenado no nó (ou seja, cada nó deve armazenar a informação sobre quantas vezes o respectivo padrão ocorreu na imagem).

**O que o programa deve fazer:**

- O programa deve ter uma fase de construção das árvores e outra de busca sobre essas árvores.
- Na construção da árvore, considera-se que a imagem binária a partir da qual os dados (padrões) encontra-se armazenado em um arquivo. O nome desse arquivo deverá ser passado como parâmetro pela linha de comandos, ao programa. Vamos considerar novamente arquivos no formato PGM, como no EP1. As dimensões da janela,  $p$  e  $q$ , deverão também ser passados como parâmetro pela linha de comandos.

Os padrões devem ser obtidos percorrendo-se a imagem no sentido raster (i.e., uma linha após a outra, de cima para baixo e cada linha da esquerda para a direita). Por exemplo, se a imagem e a janelas possuírem dimensões  $5 \times 5$  e  $3 \times 3$  respectivamente, então o número total de padrões a serem coletados é 9 (note, no entanto, que o número de padrões distintos e, portanto, o número de nós da árvore poderá ser menor, pois podem ocorrer padrões repetidos).

O programa deve ter uma opção para que o usuário possa selecionar o tipo de árvore a ser construída. Use 1 para árvore sem balanceamento, 2 para AVL e 3 para RB-Tree.

- Para cada árvore construída, deverão ser impressos:
  - Número de nós na árvore
  - Número total de padrões coletados (soma das frequências)
  - A altura (ou seja, o número de níveis de nó) da árvore
  - Número total de rotações realizadas (para as árvores balanceadas). No caso de dupla rotação, contar duas rotações.
  - O comprimento de caminho interno e externo,  $I(T)$  e  $E(T)$  respectivamente, conforme definidos em sala de aula, e respectivas médias
  - O comprimento de caminho interno ponderado  $I'(T)$  conforme definido em sala de aula, usando a frequência armazenada em cada nó, e respectiva média
  - O tempo de CPU para a construção da árvore<sup>1</sup>.
- Na fase de busca, deve-se utilizar uma segunda imagem binária e a mesma janela usada na construção das árvores. Cada padrão coletado deverá ser buscado na árvore construída.
- Para as buscas realizadas sobre cada árvore, deverão ser impressos:
  - Número de padrões buscados
  - Número de buscas com sucesso e sem sucesso
  - Número de comparações<sup>2</sup> totais no caso de buscas com sucesso e sem sucesso e respectivas médias.
  - O tempo de CPU para realizar as buscas de todos os padrões da imagem<sup>3</sup>.

**Parte opcional** (que dará direito a um bônus de no máximo 3 pontos no EP): Seu EP poderá ter opções adicionais de percorrimento da imagem de entrada na fase de construção da árvore. Além do percorrimento na ordem raster, poderá incluir opções para percorrimento na ordem anti-raster e também percorrimento aleatório (neste caso, deve-se tomar cuidado para que cada posição válida da imagem seja visitada e visitada apenas uma única vez). O percorrimento aleatório poderá ser repetido várias vezes e para aqueles valores de saída solicitados para os quais estatísticas como média e variância fazem sentido, os mesmos podem ser impressos. Esta parte opcional teria como objetivo analisar se a ordem de inserção afeta, e em caso positivo, o quanto afeta o desempenho da árvore resultante. Seu EP poderá também implementar os dois tipos de árvores balanceadas (AVL e RB-Tree).

**Parâmetros da linha de comandos:** para uniformizar a correção do EP, os dados de entrada devem ser especificados via parâmetros a serem fornecidos pela linha de comandos. Os parâmetros são os seguintes:

- nome do arquivo com a imagem para a construção das árvores
- dimensões da janela
- nome do arquivo com a imagem para a fase de buscas em árvores
- tipo da árvore a ser construída
- outros parâmetros adicionais necessários (por exemplo, ordem de percorrimento, quantidade de percorrimentos)

Quando executado sem parâmetro, o seu EP deverá descrever os parâmetros que devem ser especificados.

<sup>1</sup>Para o tempo de CPU, não contabilize o tempo de leitura da imagem, nem o de eventuais percorrimentos necessários após o término da construção da árvore.

<sup>2</sup>Uma comparação refere-se a uma comparação entre o padrão buscado e um padrão armazenado em um nó, não importando quantas operações de comparação da linguagem de programação estejam envolvidas.

<sup>3</sup>Para o tempo de CPU, não contabilize o tempo de leitura da imagem.