

Introdução

Vimos em aula quatro diferentes estratégias para busca de palavras em textos, força-bruta, KMP, Boyer-Moore e Rabin-Karp. Os três últimos métodos seguem alguma estratégia de pré-processamento para melhorar o desempenho, quando comparado à estratégia força-bruta. Entretanto, durante as aulas, não ficou claro qual deles é melhor ou se há algum que oferece desempenho claramente superior aos demais. O objetivo deste trabalho comparar experimentalmente os métodos de busca de forma a responder, com base em dados experimentais, as seguintes perguntas:

- Qual o comportamento dos métodos em função do tamanho do alfabeto?
- Qual o comportamento dos métodos em função do tamanho do padrão?

Descrição do trabalho (avaliação experimental)

Para responder a estas perguntas, sugere-se realizar diversas buscas de padrões em diversos textos, de tamanho relativamente grande. Recomenda-se variar as características do texto e do padrão (ex.: texto em português, inglês, aleatoriamente gerados, programas de uma linguagem de programação etc). Para o padrão, é interessante realizar buscas para as quais sabe-se que será mal sucedida, a fim de forçar o processamento em todo o texto. As decisões sobre o número de experimentos, os tipos de textos e suas variações farão parte dos aspectos a serem avaliados no trabalho.

Objeto de entrega e avaliação (relatório)

Para descrever os experimentos e as conclusões obtidas, um relatório deve ser escrito. Espera-se que este relatório contenha ao menos: (a) introdução ao problema e pesquisa a estudos similares; (b) descrição dos experimentos, com justificativas para as decisões tomadas; (c) apresentação dos dados encontrados, ressaltando os pontos importantes observados; (d) conclusões.

Avaliação e nota

O relatório será avaliado como resultado final do trabalho. A sua entrega deverá ser feita em arquivo com formato pdf. As seguintes características serão consideradas como critérios de avaliação: (a) clareza do texto; (b) consistência das conclusões frente aos resultados observados; e (c) abrangência dos experimentos.