Resenha: A Composite Boyer-Moore Algorithm for the String-Matching Problem

**Autor:** Zhengda Xiong   
**Instituição:** College of Computer Science and Technology, Huazhong University of Science and Technology

O problema abordado no artigo é de correspondência de string, ou "string matching", é a tarefa de encontrar ocorrências de uma string padrão dentro de outra string ou corpo de texto. Esse problema também é conhecido como correspondência exata de string, busca de string e busca de texto.

Essencialmente, o problema é: tendo um padrão (P) e um texto (T), ambos compostos de caracteres de um alfabeto finito (Σ), como encontrar todas as posições no texto onde uma cópia exata do padrão começa?

Os autores se deparam com tal problema, fundamental em ciência da computação, biologia molecular e engenharia genética. Reconhecendo a necessidade de algoritmos eficientes para encontrar todas as ocorrências de um padrão em um texto, especialmente em grandes conjuntos de dados. Porém, o artigo, ao analisar algoritmos para string matching como BF, KMP e BM, notasse que ainda não há a eficiência desejada.

É proposto, então, uma abordagem composta para o problema de correspondência de strings, baseada no conhecido algoritmo Boyer-Moore (BM). Reconhecendo que o algoritmo Boyer-Moore sozinho pode não ser eficiente em todos os casos, foi desenvolvido uma técnica que combina o Boyer-Moore com outras estratégias para aumentar sua eficiência, o Composite Boyer-Moore (CBM). Nele a questão-chave é como utilizar as informações de comparação obtidas na iteração anterior, usando uma tabela bidimensional que indica a distância do “shift” do padrão (P).

O artigo demonstra que essa abordagem composta supera o desempenho de outras técnicas conhecidas, apresentando análises teóricas e experimentais. Esses resultados destacam a eficácia do algoritmo proposto o comparando com o BM, onde é visto as limitações do BM em alfabetos grandes, principalmente em “binary matching”, e, simultaneamente, mostrando a eficiência e constância do CBM.  Contribuindo significativamente para o campo da pesquisa em correspondência de strings.