

Implementação de uma Estrutura Trie para Verificação e Sugestão de Palavras em Processamento de Texto

Davi Lopes Lemos¹,
Heduardo Witkoski Barcelos da Rocha¹

¹Campus Bagé - Universidade Federal do Pampa (UNIPAMPA)
96.413-172 – Bagé – RS – Brazil

{davilemos, heduardorocha}.aluno@unipampa.edu.br

1. Introdução

A implementação de estruturas eficientes para busca e organização de dados é um tema amplamente abordado na área de Algoritmos e Estruturas de Dados. Entre as soluções existentes, a árvore Trie destaca-se como uma estrutura de dados hierárquica capaz de otimizar operações de busca e sugestão em grandes conjuntos de palavras. Conforme descrito por Sedgewick and Wayne [2011], a Trie é amplamente utilizada para o armazenamento e a busca eficiente em contextos onde a consulta rápida a dicionários ou listas de palavras é essencial.

Neste trabalho, a estrutura Trie é explorada em dois contextos principais: a verificação de palavras válidas da língua inglesa e a implementação de um sistema de autocompletar. A verificação de palavras consiste em identificar se palavras digitadas pelo usuário pertencem a uma lista de palavras pré-definidas, utilizando como base um arquivo obtido a partir de fontes públicas que reúnem extensas coleções de palavras válidas da língua inglesa. Por outro lado, o sistema de autocompletar visa sugerir alternativas de palavras a partir de um fragmento fornecido, oferecendo ao usuário uma experiência de interação mais ágil e eficiente.

A implementação proposta tem como foco a demonstração da aplicabilidade da estrutura Trie na resolução de problemas que envolvem buscas eficientes e sugestões de palavras. A solução apresentada é fundamentada em requisitos bem definidos e estruturada de forma a evidenciar a relevância e a eficiência desta abordagem para o domínio escolhido. Além disso, o estudo realizado está alinhado às diretrizes do componente curricular, consolidando conceitos teóricos e práticos relacionados à aplicação de estruturas de árvores em soluções computacionais.

1.1. Domínio da problemática

O domínio do problema abordado neste trabalho envolve a necessidade de mecanismos eficientes para verificação e sugestão de palavras em aplicações relacionadas ao processamento de texto. Com o avanço das tecnologias digitais e a popularização de dispositivos que utilizam sistemas de autocompletar e correção ortográfica, a busca eficiente em grandes volumes de dados tornou-se um requisito essencial. Dessa forma, é fundamental contar com estruturas de dados capazes de reduzir o tempo de busca e otimizar o desempenho de aplicações interativas.

Sistemas de verificação ortográfica precisam validar, em tempo real, se palavras digitadas pelos usuários estão presentes em bases de dados predefinidas. Arquivos provenientes de fontes públicas oferecem listas robustas com milhares de palavras válidas da

língua inglesa. No entanto, métodos de busca lineares tornam-se ineficientes à medida que o volume de dados cresce, reforçando a necessidade de utilizar estruturas de dados como a Trie para garantir eficiência na operação de busca e validação de palavras.

Além disso, sistemas de autocompletar são amplamente utilizados em aplicações modernas, como editores de texto, mecanismos de busca e aplicativos de mensagens. Essas aplicações dependem de algoritmos rápidos e precisos que, a partir de um fragmento fornecido pelo usuário, consigam sugerir palavras relevantes de forma eficiente. A estrutura Trie é especialmente adequada para essa finalidade, uma vez que permite a busca por prefixos de maneira hierárquica, garantindo rapidez e precisão na geração de sugestões.

2. Metodologia

A implementação foi realizada em Java, utilizando os princípios da orientação a objetos. A solução envolveu a construção de uma estrutura Trie para operações de inserção, busca e sugestão de palavras, com base em uma lista pré-processada proveniente de fontes públicas. A linguagem Java foi utilizada devido à sua robustez e ao suporte a estruturas de dados eficientes, o que permitiu a criação de um programa modular e escalável.

Durante o desenvolvimento, foram implementadas técnicas para leitura e processamento de arquivos, permitindo a carga eficiente dos dados na estrutura Trie. A estrutura foi projetada para garantir rapidez nas operações de validação de palavras e sugestões de alternativas com base em fragmentos fornecidos pelo usuário. Sua corretude foi validada em um ambiente que utiliza o sistema operacional Ubuntu 24.04 LTS, processador AMD Ryzen 7 5700U e 12 GB de memória DDR4.

3. Resultados

A implementação da estrutura Trie demonstrou ser uma alternativa válida e eficiente para as operações de verificação e sugestão de palavras. Através dos testes realizados, foi possível observar que a solução atendeu aos requisitos de rapidez e precisão nas buscas, oferecendo resultados consistentes, mesmo neste cenário com grande volume de dados. As operações de inserção, validação e sugestão de palavras foram realizadas de maneira ágil, comprovando a eficiência da estrutura no contexto proposto, conforme ilustra a Figura 1.

```
Dicionário carregado com sucesso.
Digite uma palavra para verificar (ou 'sair' para encerrar): Hello
A palavra 'Hello' está correta.
Digite uma palavra para verificar (ou 'sair' para encerrar): World
A palavra 'World' está correta.
Digite uma palavra para verificar (ou 'sair' para encerrar): Gir
A palavra 'Gir' não está no dicionário.
Sugestões:
- girard
- girardi
- girardo
- girasol
```

Figura 1. Exemplo de utilização da solução proposta

4. Conclusão

A implementação da estrutura Trie mostrou-se uma solução eficaz para os problemas de verificação e sugestão de palavras, atendendo bem às demandas de rapidez e precisão em

aplicações de processamento de texto. Através de uma abordagem simples, porém robusta, foi possível construir um sistema que lida de forma eficiente com grandes volumes de dados, proporcionando resultados rápidos tanto na validação de palavras quanto na geração de sugestões. A escolha da linguagem Java, com seu suporte a estruturas de dados eficientes, permitiu a criação de uma solução escalável, capaz de realizar as operações de maneira ágil.

Este trabalho destaca a importância de utilizar a estrutura de dados adequada em sistemas que exigem interação rápida com grandes volumes de informações. A Trie se mostrou uma ferramenta poderosa, ideal para otimizar buscas e sugestões em contextos como verificação ortográfica e autocompletar, comuns em muitas aplicações modernas. A solução proposta, além de cumprir os requisitos do estudo, revela o potencial da Trie para resolver desafios práticos em computação de forma eficiente.

Referências

Sedgewick, R. and Wayne, K. (2011). *Algorithms*. Addison-Wesley Professional, Boston, MA, 4th edition.