FGV/EMAp

Final Project - Inverted Index and Comparative Analysis

Arthur Rabello Oliveira[1], Gabrielle Mascarelo, Eliane Moreira, Nícolas Spaniol, Gabriel Carneiro

Abstract

We present the implementation and comparative analysis of three fundamental data structures for indexing and searching textual documents: the Binary Search Tree (BST), the AVL Tree, and the Red-Black Tree (RBT). Each structure was implemented with its core operations, including insertion and search. Unit tests were developed to validate the correctness and performance of these implementations. We also provide a further comprehensive comparative study of the three trees based on their time complexity, balancing efficiency, and suitability for document indexing. The results demonstrate the trade-offs between implementation complexity and query performance, offering insights into the practical considerations for choosing appropriate search tree structures in information retrieval systems.

Contents

[1]Escola de Matemática Aplicada, Fundação Getúlio Vargas (FGV/EMAp), email: arthur.oliveira.1@fgv.edu.br

5.1. The Experiment

5.2. Memory Usage

5.3. Time Complexity

6. Conclusion

6.1. Summary of Findings

7. Source code

(repository)

8. Task Division (Required by the professor)

8.1. Arthur Rabello Oliveira

Contributed with:

- Keeping the repository civilized (main-protecting rules, enforcing code reviews)
- Writing and documenting the `Makefile` for building the project
- Writing the `README.md` and `report.typ`
- Writing and documenting functions for the classic BST in `bst.cpp`

8.2. Gabrielle Mascarelo

Contributed with:

- Writing and documenting functions to read files in `data.cpp`
- Structuring statistics in the CLI

8.3. Eliane Moreira

Contributed with:

- Testing all function related to the BST
- Writing and documenting functions for `tree_utils.cpp`
- Fixing bugs in `data.h`

8.4. Nícolas Spaniol

Contributed with:

- Code reviews and suggestions for improvements in the codebase.
- Built from scratch the JavaScript visualization of all trees

8.5. Gabriel Carneiro

Contributed with:

- Writing and documenting functions for the classic BST
- Writing and documenting functions for the AVL Tree
- Testing functions for `tree_utils.cpp`