Projeto de Laboratórios de Informática III
Grupo 19

André Sousa a78322 Carlos Pedrosa a77320 David Sousa a78938
Manuel Sousa a78869

26 de Abril de 2017

Resumo

Este projeto tem como principal objetivo a criação de um sitema que permita analisar os artigos presentes em backups da Wikipedia, realizados em diferentes meses, e extrair informação útil para esse período de tempo como, por exemplo, o número de revisões, o número de novos artigos, etc.

Assim, este relatório pretende sumarizar todos os esforços efetuados para alcancar o objetivo proposto. Nesse sentido, serão apresentadas as soluções feitas nas diversas tarefas propostas pelos docentes.

Conteúdo

1 Introdução

Este trabalho foi-nos proposto no âmbito da unidade curricular de Laboratórios de Informática III e tem como principal objetivo fazer com que haja, por parte dos alunos, um maior contacto com a linguagem de programação impertativa C bem como a sensibilização destes para a escolha de código rapido e eficiente quando se trata do manuseamento de grandes quantidades de dados.

Nesse sentido, foi-nos apresentado um projeto dividido em duas tarefas fundamentais, no qual o fim único da sua realização é a consulta rápida dos dados, usando para isso certas querys que foram fornecidas pelos professores. Assim, neste relatório, vamos descrever todos os passos dados para alcancar o objetivo proposto, bem como os problemas que foram surgindo aquando a realização desses mesmos passos.

Em suma, pretendemos falar do desenvolvimento e da concepção do código, explicando, deste modo, a forma de como efetuamos cada uma das tarefas sugeridas pelos docentes.

2 Descrição do Problema

O objetivo geral do projeto apresentado é o desenvolvimento de um sistema de pesquisa rápido. Para esse efeito, os problemas foram abordados conforme o enunciado apresentado.

Efetivamente, num primeiro momento,deparamo-nos com a biblioteca xml, pelo que o primeiro problema que tivemos foi o manuseamento desta.

Por último, os snapshots disponibilizados pelos professores são de tamanho relativamente elevado. Nesse sentido, é necessária uma estrutura rápida e concisa que permita um acesso aos dados quase imediato. Sendo assim, foi outra das dificuldades que tivemos de ultrapassar.

3 Concepção da Solução

Ao longo do trabalho efetuado fomos testando as nossas soluções no site disponibilizado pelo professor. Assim, desenvolvemos o código necessário para dar resposta a todas as interrogações que foram propostas.

3.1 Estruturas de Dados

As *estruturas de dados* são um modo de armazenamento e organização de dados, de modo, que podem ser usadas de forma eficiente, facilitando assim a busca e modificação.

Nesse sentido, como era necessária uma estrutura de dados relativamente eficiente optamos por utilizar uma biblioteca do sistema, a *Glib*. Assim, dentro de uma grande panóplia de estruturas oferecida por esta biblioteca usamos a *GTree*(balenced binary tree), uma vez que, considerarmos ser a estrutura que mais se adequava ao trabalho proposto. Deste modo, esta é em tudo semelhante a uma árvore balanceada, tal como o proprio nome indica,pelo que, consideramos relevante a utilização desta devido à sua forma de implementação rapida no que trata ao acesso dos dados.

3.2 Dificuladades na concepção da solução

No decorrer do trabalho deparamo-nos com diversas dificuldades, sendo que, na sua grande maioria foram ultrapassadas. No entanto, apesar de tudo, tivemos alguns problemas que apenas foram parcialmente resolvidos. Assim, mesmo usando a estrutura de dados que nos pareceu ser a mais adequada, continuamos a obter tempos aquém do que seria esperado. Tal problema, pode dever-se não só à estrutura que usamos mas também ao modo como as funções foram implementadas. No caso da função load, sendo esta a que tem um maior impacto em todo o programa, usamos listas ligadas que podem influenciar negativamente o tempo de execução do ficheiro quando a sua utilização se torna demasiado elevada. De modo a resolver estes problemas pensamos no uso de uma tabela de Hash, também ela, implementada na *Glib*. Contudo, a escolha desta estrutura era um pouco arriscada pois tal implementação requeria uma função de Hash bastante eficiente.

4 Conclusões

De acordo com o objetivo do projeto, sendo este o acesso rápido a um conjunto elevado de dados, podemos concluir que este foi razoavelmente concluido.

De facto, apesar das nossas tentativas, não nos foi possível obter um tempo de execução menor. Consideramos, assim, que o grande objetivo foi alcançado. Efetivamente, e apesar de tudo, o projeto no que diz respeito à primeira fase foi satisfatoriamente resolvido pelo que o código necessário para o funcionamento das queries foi desenvolvido.

Em jeito de conclusão, o presente trabalho serviu para um aprimoramento das nossas aptidões relativamente à linguagem de programação imperativa C. Nesse sentido, este trabalho formatou a nossa maneira de pensar aquando da resolução de um problema, tornando-nos mais objetivos e racionais.