**Recuperação de Informação**

Trabalho prático 1

Ewerton S. Santos

Departamento de Ciência da Computação

Universidade Federal de Minas Gerais

Belo Horizonte, MG - Brasil

[ewerton\_dc@hotmail.com](mailto:ewerton_dc@hotmail.com)

Matricula: 2016058140

Belo Horizonte, MG

Dezembro, 2020

1. **Introdução**

Adicionar referencias !!!!

Um motor de busca na web moderna possui três componentes principais: um coletor de páginas (crawler); um ***ALGUAM COISA QUE NÃO LEMBRO***; um processador de consultas. Como primeira etapa do processo de recuperação de informação faz-se necessária a obtenção dos documentos web, no casos os arquivos HTML de cada página, para que estas sejam armazenadas e processadas para as consultas.

No entanto, no primeiro momento temos que partir de um conjunto de sítios web relevantes, os quais vão nos levar as demais paginas do mesmo e a paginas de outros servidores. Dessa forma, em uma pagina obtemos novas paginas para serem coletadas, sendo um processo incremental e repetido algumas vezes para atualizar o conteúdo.

1. **Desenvolvimento**

Para desenvolver o nosso coletor utilizou-se a linguagem C++, sistema operacional Ubuntu 18.04 e os compentes gratuitos da biblioteca Chilkat®, mais especificamente o modulo Spider.

Para realizar o processo de coleta fora criado um modulo que recebe um numero inteiro **n** e um string contendo a url do domínio a ser coletado. O meotod spider.Initialize define recebe a url a ser raspada. O meotodo spider.CrawNext caminha pelas lista de URL recebida, fazendo uma requisição para nova url. Os métodos lasHTMLtitle retorna o titulo da url requisitada e lastURL retorna a url para qual foir realizada a requisição.

No intervalor entre cada requisição é contabilizado o tempo de inicio e o tempo de fim, o tempo de cada requisição é somado a uma variável double, ao fim do processo é calculada a media do tempo de execução para cada request.

1. **Conclusão**

Modelo de texto já formatado para realizar a criação dos parágrafos. Apenas um exemplo formato do texto, use este modelo com os formatos, poupa muito do tempo de formatar um texto.

1. **Referencias**

Modelo de texto já formatado para realizar a criação dos parágrafos. Apenas um exemplo formato do texto, use este modelo com os formatos, poupa muito do tempo de formatar um texto.