

Universidade de Brasília
Departamento de Ciência da Computação
Disciplina: Projeto e Análise de Algoritmos
Código da Disciplina: 117536

Projeto e análise de algoritmos 1/2021

Projeto de Disciplina:

Fazer uma search engine, sendo dividida nas seguintes partes:

- a) webcrawler: responsável por acessar as páginas web e salvar as páginas e respectivas urls. Não precisa ser implementado neste trabalho. Pode ser usado um programa ou biblioteca separada
- b) armazenar as páginas e colocar em um formato que pode ser acessado. Deve ser totalmente implementado pelo grupo
- c) indexar as páginas e fazer a pesquisa de forma rápida. Deve ser totalmente implementado pelo grupo.
- d) fazer um servidor web que permita usar o serviço web de busca e mostrar os resultados. Podem ser usados programas específicos para servidores.

A search engine deve permitir as seguintes pesquisas:

- a) pesquisa por palavras palavras ex. unb notícias : vai procurar por páginas que contém as palavras unb e notícias. Pode ser usado unb AND notícias com o mesmo significado
- b) pesquisa sem uma palavra ex. unb –vestibular : vai procurar por páginas que contém as palavras unb mas não a palavra vestibular
- c) pesquisa por string ex. “notícias da unb” : vai procurar por páginas que contém a string “notícias da unb”
- d) pesquisa por um termo ou outro. Ex “unb” OR “universidade de Brasília” : pesquisa por páginas que tem o termo “unb” ou “universidade de Brasília” ou os dois.
- e) a pesquisa pode ter qualquer número de palavras, strings, operadores AND e OR e parênteses. Ex (unb OR “universidade de Brasília”) AND “notícias atuais”

f) o resultado deve vir em páginas web onde os resultados são mostrados com: url, título da página, trecho do texto onde a procura aparece, link para o usuário clicar e ir para a página.

Os algoritmos utilizados devem ser eficientes permitindo a pesquisa rápida.

Nesta fase inicial é apenas para a definição dos grupos e do que vai ser utilizado no projeto.

Deve ser enviado:

- 1) Os integrantes do grupo. São no máximo 7.
- 2) Dizer quais linguagens deverão ser utilizadas
- 3) dizer quais programas serão utilizados para o webcrawler, servidor, etc.
- 4) Dizer que bibliotecas, frameworks serão utilizados. Deverá ser verificado com o professor e deverá seguir os seguintes critérios:
 - 4.1) open source
 - 4.2) Licenças : GPL, LGPL, Apache License 2.0, BSD, MIT ou Mozilla
- 5) Um documento “.pdf“ contendo descrevendo quais algoritmos se pretende utilizar e a análise da complexidade dos algoritmos.

Pela tarefa na página da disciplina no aprender3.unb.br

Apenas um integrante do grupo deve enviar o trabalho. O nome do trabalho deve ser algo como:

PAA_Jose_12345_Proj.zip

Com o primeiro nome e matricula de quem envia pelo grupo.

Deve ser enviada pelo aprender3.unb.br:

Data de Entrega 3/10/21 até as 23:55