

**PROGRAMAÇÃO AVANÇADA**

**Joanã Costa 180 221 109**

**Daniel Cordeiro 190 221 101**

**WebCrawler**



Índice

[Introdução 3](#_Toc25538930)

[Pseudocódigo 3](#_Toc25538931)

[Modo Automático 3](#_Toc25538932)

[Número de Páginas Visitadas 3](#_Toc25538933)

[Profundidade 4](#_Toc25538934)

[Modo Iterativo 4](#_Toc25538935)

[Mockup da interface 5](#_Toc25538936)

# Introdução

Este projeto tem como objetivo a utilização da TAD ***Graph*** e demais ***Queue*** e ***Stack*** na resolução do problema de percorrer páginas *web* e extrair demais páginas associadas. Sendo assim será desenvolvido um *WebCrawler* que fará essa tarefa através dos algoritmos ***breadth-fisrt***, ***depth-first*** e eventualmente o algoritmo *Djisktra*.

**Rafael Trindade - 160210033**

**Vítor Gomes - 160210050**

# Pseudocódigo

Nesta parte iremos mostrar através de pseudocódigo como o nosso algoritmo está escrito.

# Modo Automático

## Número de Páginas Visitadas

BFS (webPageRoot)

Cria uma variavel para contar o número de páginas visitadas

Criar uma lista iterável que vai ser retornada no método

Criar uma fila para guardar as páginas a visitar

Verificar se o numero de páginas pretendidas já foi atingida

Se sim -> retornar a lista iterável de resultado

Verificar se a webPageRoot não existe já no webCrawler:

Se não -> Inserir no webCrawler a webPageRoot como Vertex

Adicionar a webPageRoot como primeira página a ser visitada

Adicionar á lista iterável de resultado

Incrementar a variavel das páginas visitadas

Enquanto a fila não está vazia faça:

Retirar da fila de páginas a visitar e guardar a primeira webPage da fila numa variável

Fazer print da página visitada

Guardar numa fila todos os links incidents a esta webPage visitada

Iterar sobre a fila obtida:

Verificar se o numero de páginas pretendidas já foi atingida:

Se sim -> retornar a lista iterável de resultado

Inserir um novo vértice com uma nova webPage com o link atual iterado

Adicionar esta webPage á lista iterável de resultado

Adicionar esta webPage á fila de páginas a visitar

Fazer print do link incidente

Inserir um novo edge com vértice inicial a webPage a ser visitada e como vértice final a nova WebPage incidente

Incrementar a váriavel das páginas visitadas

Retornar lista resultado

## Profundidade

Mesmo tentando fazer o BFS com um critério de nível de profundidade, não conseguimos implementar o pseudocódigo.

# Modo Iterativo

Neste modo será possível o utilizador gerar o seu próprio grafo através de cliques nas respetivas páginas.

BFS ITERATIVO (Graph, vértice)

Marque vértice como visitado

Marque vértice como último a restaurar

Coloque o vértice na fila

Enquanto a fila não estiver vazia faça

* Seja v o primeiro vértice da fila
* Para cada vértice w adjacente a v faça
  + Se w não está marcado
  + Então
    - Marque w
    - Insira no final da fila
* Se último a restaurar
  + Para cada vértice w adjacente a v faça
    - Se não está marcado
    - Então
      * Marque w
      * Insira no final da fila

# Mockup da interface

Este *mockup* é uma pequena demonstração do que será a aplicação física em si.

