

 <p>INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA PIAUÍ</p>	<p><b>INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ</b></p> <p><b>Curso: ADS</b></p> <p><b>Disciplina: Programação para a Internet</b></p> <p><b>Professor: Ely</b></p>
--	--

### Exercício sobre buscadores – Parte I

- 1) Analise e estude o conteúdo das 5 páginas sobre filmes de ficção fornecidas. As principais características das páginas:
  - a) Elas possuem links entre si, sendo que uma delas possui um “autoreferência”;
  - b) Possuem termos e palavras-chave no cabeçalho, tags h1, h2, parágrafos e links;
  - c) Segue uma tabela que mostra a ligação entre as páginas:

Página de Origem	Quem aponta para ela	Quantidade de Links que apontam
blade_runner.html	Blade runner (autoreferência), Interestellar, Matrix, Mochileiro	4
duna.html	Matrix	1
matrix.html	Blade runner, Interestellar, Mochileiro	3
interestellar.html	Blade runner, duna, Matrix, mochileiro	4
mochileiro.html	Blade runner, duna, matrix	3

- 2) Implemente um buscador com os seguintes Critérios de Ranqueamento:
  - a) Autoridade das Páginas: Determinada pela quantidade de links recebidos de outras páginas:  
Pontos: Cada link recebido vale +20 pontos.
  - b) Quantidade dos Termos Buscados: Frequência com que os termos buscados aparecem no código-fonte da página:  
  
Pontos: Cada ocorrência do termo buscado vale +5 pontos. Caso não existam ocorrências, a página não deve ser listada
  - c) Uso das Tags (head, h1, h2, p) para Relevância:

Pontos: Uso de termos buscados em title e meta tags (+20 pontos cada), h1 (+15 pontos cada ocorrência), h2 (+10 pontos cada), p (+5 pontos cada), a (+2 pontos)

d) Penalização por Autoreferência:

Pontos: Cada autoreferência resulta em uma penalidade de -20 pontos.

e) Frescor do Conteúdo: Avaliado pela data de publicação da página.

Pontos: Páginas publicadas no ano corrente recebem +30 pontos, com uma redução de -5 pontos para cada ano anterior.

3) Use os seguintes critérios de desempate:

- Maior quantidade de termos buscados no corpo do texto;
- Maior frescor do conteúdo (datas mais recentes);
- Maior número de links recebidos.

4) Considere ainda as seguintes recomendações:

- O seu sistema deve a partir de uma página inicial baixar o conteúdo para posterior aplicação de indexação;
- Verifique seus links armazenando-os em alguma estrutura de dados;
- Para cada link na página, baixe a página referenciada e reexecute o passo a)
- Mantenha em um meio de persistência uma estrutura de dados a ser consultada para que uma mesma página não seja visitada/baixada duas vezes;

5) Implemente através de um arquivo de configuração ou painel de controle para a personalização para as pontuações do item 2.

6) Implemente alguns casos de teste, como por exemplo, uma busca pela palavra Matrix:

Pesquisa pelo termo matrix							
Página	Autoridade (+10/link)	Frequência do termo (+5/ocorrência)	Uso em tags (+20/+15/+10/+5/+2)	Autoreferência (-20)	Frescor do conteúdo(+30/-5/ano)	Total	Deve ser exibida
matrix.html	30	45	85	0	-30	130	Sim
interestelar.html	40	20	22	0	-15	67	Sim
mochileiro.html	30	5	5	0	0	40	Sim
duna.html	10	0	0	0	20	30	Não
blade_runner.html	40	25	22	-20	-40	27	Sim

<https://docs.google.com/spreadsheets/d/1JKAdHJazf4Wv8bpZQ8HIJxIDrg2to4iTOqtcoFzQdKc/edit?usp=sharing>

7) Exiba os resultados da busca da forma mais amigável que conseguir e exiba, se solicitado os critérios de ranqueamento com uma tabela semelhante à demonstrada acima.