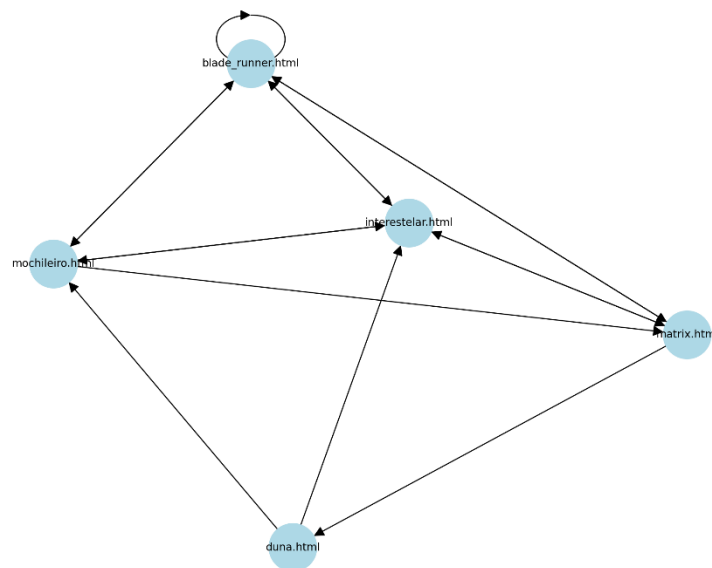
 <p>INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA PIAUÍ</p>	<p>INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ</p> <p>Curso: ADS</p> <p>Disciplina: Programação para a Internet</p> <p>Professor: Ely</p>
--	--

Exercício sobre buscadores

- 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

Grafo de Ligações entre as Páginas (Incluindo Auto-Referências)



- 2) Hospede as 5 páginas no githubpages de forma que seja possível navegar entre os seus links.
- 3) Crie uma implementação iniciando como um crawler com os seguintes passos:
 - a. Receba um link para uma das páginas iniciais;
 - b. Baixe a página e extraia o seu conteúdo e seus links;
 - c. Para cada link, repita os passos a e b até que todas as páginas tenham sido baixadas;
 - d. Mantenha em um meio de persistência uma estrutura de dados a ser consultada para que uma mesma página não seja visitada/baixada duas vezes.
- 4) De posse das páginas já baixadas, 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 +10 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 +10 pontos.
 - c) Penalização por Autoreferência:
Pontos: uma penalidade de -15 pontos.
- 5) Use os seguintes critérios de desempate:
 - a. Maior número de links recebidos;
 - b. Maior quantidade de termos buscados no corpo do texto;

c. Autorreferência.

6) Casos de teste:

Caso	Termo Buscado	Onde aparecem
1	Matrix	Aparece em Blade Runner, Interestelar, Matrix, Mochileiro
2	Ficção Científica	Tema em Blade Runner, Matrix, Mochileiro, Interestelar
3	Realidade	Tema em Blade Runner, Matrix, Interestelar
4	Universo	Presente em Duna, Mochileiro
5	Viagem	Presente em Mochileiro, Interestelar

a) Busca pelo termo Matrix

Posição	Página	Ocorrências (+5)	Links Recebidos (+10)	Autoreferência (-15)	Total
1	matrix.html	8×5 = 40		30	0
2	interestelar.html	3×5 = 15		40	0
3	mochileiro.html	2×5 = 10		30	0
4	blade_runner.html	2×5 = 10		40	-15

b) Busca pelo termo ficção científica:

Posição	Página	Ocorrências (+5)	Links Recebidos (+10)	Autoreferência (-15)	Total
1	interestelar.html	1×5 = 5	40	0	45
2	matrix.html	3×5 = 15	30	0	45
3	mochileiro.html	2×5 = 10	30	0	40
4	blade_runner.html	2×5 = 10	40	-15	35

c) Busca pelo termo realidade:

Posição	Página	Ocorrências (+5)	Links Recebidos (+10)	Autoreferência (-15)	Total
1	interestelar.html	2×5 = 10	40	0	50
2	matrix.html	2×5 = 10	30	0	40
3	blade_runner.html	1×5 = 5	40	-15	30

d) Busca pelo termo universo:

Posição	Página	Ocorrências (+5)	Links Recebidos (+10)	Autoreferência (-15)	Total
1	mochileiro.html	2×5 = 10	30	0	40
2	duna.html	1×5 = 5	10	0	15

e) Busca pelo termo viagem:

Posição	Página	Ocorrências (+5)	Links Recebidos (+10)	Autoreferência (-15)	Total
1	interestelar.html	1×5 = 5	40	0	40
2	mochileiro.html	2×5 = 10	30	0	30

- 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.
- 8) Crie uma nova página HTML com os mesmos padrões e adicione mais dois casos de teste com novos termos de busca. Valide seu resultado repassando a página e termos de buscas à outra equipe e apresente o trabalho com sua equipe no dia especificado.

Sugestão de leitura:

- <https://moz.com/beginners-guide-to-seo/how-search-engines-operate>
- <https://axios-http.com/docs/intro>
- <https://cheerio.js.org/>

Exemplo de download de página em JavaScript:

```
const axios = require('axios');
const cheerio = require('cheerio');

async function crawlPagina(url) {
  try {
    const resposta = await axios.get(url);
    const $ = cheerio.load(resposta.data);

    const links = [];
    $('a').each((i, elemento) => {
      const texto = $(elemento).text();
      const href = $(elemento).attr('href');
      if (href) {
        links.push({ texto, href });
      }
    });
  }
}
```

```
    console.log("Links encontrados:", links);
  } catch (erro) {
    console.error("Erro ao acessar a página:", erro.message);
  }
}

// Exemplo de uso:
crawlPagina('https://developer.mozilla.org');
```