

Teste técnico: Backend Software Developer

Como parte do processo seletivo para a posição de Backend Software Developer, você deve desenvolver uma aplicação de *web crawling*. Mas não se preocupe: nenhum conhecimento específico de *web crawling* é necessário além de noções básicas de HTML. O objetivo da aplicação é navegar por um website em busca de um termo (sequência de caracteres) fornecido pelo usuário, seguindo links a partir da página inicial determinada e avaliando o conteúdo de cada página visitada.

Este exercício permitirá que você demonstre sua capacidade de resolver problemas e seus conhecimentos e habilidades de desenvolvimento de software. Para focar nisso, o uso de *frameworks* e bibliotecas externas é limitado — você deve explorar o poder da API padrão da plataforma Java para atingir seu objetivo.

Funcionamento

A interação do usuário com a aplicação é limitada ao fornecimento de argumentos de execução por meio de variáveis de ambiente pela linha de comando (ver seções **Variáveis de ambiente** e **Compilação e execução**, abaixo). As variáveis de ambiente fornecidas determinam a URL base do website em que a busca deve ser feita, o termo a ser buscado e, opcionalmente, o número máximo de resultados que devem ser obtidos antes da busca ser interrompida.

A busca deve iniciar pela URL base e seguir links (absolutos e relativos) em elementos *anchor* das páginas visitadas se e somente se eles possuírem a mesma URL base que é especificada para a inicialização da aplicação. Exemplo: se a URL base foi definida como `http://ex.com/site/`, um link para `http://ex.com/site/pagina.html` deve ser visitado; já um link para `http://ex.com/blog/` não deve ser visitado.

Uma URL visitada é considerada um resultado da busca se e somente se sua página contiver o termo buscado, de forma *case-insensitive*, em qualquer parte do conteúdo HTML (dentro ou fora de *tags*). Os resultados encontrados devem ser escritos no *standard output*, com uma única URL absoluta por linha, de acordo com o formato abaixo (onde o texto `URL` deve ser substituído pela URL do resultado). A ordem de apresentação dos resultados é irrelevante, mas não deve haver repetições.

```
Result found: URL
```

Cada linha de resultado deve iniciar com o texto `"Result found:"`. Linhas escritas no *standard output* fora desse padrão serão ignoradas (o que permite *logs* de processamento, por exemplo).

Pode-se assumir (sem validações específicas por parte da aplicação) que todas as páginas que devem ser visitadas são formadas por conteúdo HTML válido, e que esse conteúdo não será modificado no servidor que hospeda o website durante a execução da aplicação.

Variáveis de ambiente

- **BASE_URL:** Definição obrigatória. Determina a URL base do website em que a busca deve ser feita pela aplicação. Restrições: o valor deve conter uma URL (HTTP ou HTTPS) válida e absoluta de acordo com a implementação da classe `java.net.URI`.
- **KEYWORD:** Definição obrigatória. Determina o termo (sequência de caracteres) a ser buscado no conteúdo das páginas visitadas pela aplicação. Restrições: o valor deve conter no mínimo 4 e no máximo 32 caracteres, e deve conter apenas caracteres alfanuméricos.
- **MAX_RESULTS:** Definição opcional. Determina o número máximo de resultados que devem ser retornados por uma busca. Restrições: O valor deve representar um número inteiro igual a -1 (indicando limite não definido, ou seja, busca ilimitada) ou maior do que 0. Quando o valor não é especificado, ou é especificado um valor inválido, a aplicação deve assumir o valor *default* -1. Quando uma busca atinge o limite de resultados, ela deve ser concluída, deixando de visitar novos links encontrados. Buscas com limite máximo definido podem retornar resultados diferentes de acordo com a estratégia adotada para navegação entre links; diferenças desse tipo são aceitáveis e não prejudicarão sua avaliação.

A aplicação é responsável pela validação dos valores de variáveis de ambiente durante sua inicialização. Caso valores inválidos sejam utilizados para uma variável obrigatória, a execução deve ser interrompida.

Tecnologias e estrutura do projeto

Você deve desenvolver sua aplicação utilizando Java, Apache Maven e Docker, com a IDE ou editor de sua preferência. As seguintes bibliotecas podem ser utilizadas: [JUnit 5](#), [Hamcrest](#) e [Mockito](#).

Junto com este documento, você recebeu um arquivo compactado que contém uma estrutura parcial do projeto que irá desenvolver. A estrutura segue o padrão estabelecido pelo Apache Maven. Os

arquivos marcados na listagem abaixo em negrito com asteriscos **não podem ser modificados ou movidos**; qualquer modificação levará à sua **desclassificação** do processo seletivo. Os demais arquivos e diretórios podem ser modificados à vontade.

```
/src
  /main
  /test
Dockerfile *
pom.xml *
```

Visto que o arquivo `pom.xml` não deve ser modificado, o uso de *frameworks* e bibliotecas externas é limitado, como já foi mencionado. Todas as dependências que podem ser utilizadas pela sua aplicação já estão listadas no arquivo fornecido.

Compilação e execução

A partir do diretório raiz do projeto, os seguintes comandos, executados em sequência, devem fazer a compilação e execução da aplicação:

```
docker build . -t axreng/backend
docker run -e BASE_URL=http://hiring.axreng.com/ -e KEYWORD=four --rm axreng/backend
```

Recomendamos utilizar a versão 19.03.6 ou posterior do Docker Engine.

Os comandos descritos, bem como a passagem de argumentos e o registro de resultados (ver seção **Funcionamento**, acima), **devem ser seguidos à risca**, visto que os testes serão executados de forma automatizada e qualquer divergência poderá impactar seus resultados.

Durante o desenvolvimento, incluir no comando `docker run` o parâmetro `-v "$HOME/.m2":/root/.m2` (ou o equivalente para o seu repositório Maven local) pode reduzir significativamente o tempo de download de dependências.

Critérios de avaliação

A solução que você desenvolver e entregar será avaliada de acordo com as seguintes etapas:

- 1. Requisitos funcionais.** Nessa etapa são executados scripts previamente desenvolvidos que fazem a compilação e a execução de sua aplicação. O funcionamento da busca é validado com testes em múltiplos websites simples criados especificamente para esse propósito. Os resultados obtidos são comparados com os resultados esperados, e qualquer divergência leva à interrupção da avaliação, sem o prosseguimento às etapas seguintes.
- 2. Qualidade interna.** Se forem atendidos plenamente os requisitos de funcionamento, a qualidade interna do seu código será avaliada por desenvolvedores experientes. É esperada a aplicação de boas práticas de desenvolvimento em geral, de acordo com padrões e referências amplamente disseminados: código limpo, princípios e padrões de *design* orientado a objetos (incluindo SOLID), etc. **Atenção:** uma solução com qualidade interna deficiente poderá ser eliminada mesmo atendendo integralmente aos requisitos funcionais.
- 3. Desempenho.** Como última etapa de avaliação, sua aplicação poderá ter seu desempenho mensurado em relação a uso de memória e tempo de processamento em cenários de teste com dificuldade crescente (como, por exemplo, execução em ambientes com restrições rígidas de memória). Essa etapa não é eliminatória.

Auxílio ao desenvolvimento

Disponibilizamos como recurso que pode auxiliar o desenvolvimento da sua solução um website simples onde podem ser feitas buscas: <http://hiring.axreng.com/>

Junto com este documento, você também recebeu um arquivo `results_four.txt` contendo um exemplo de *output* com os resultados esperados para a busca pelo termo “four” no website fornecido como exemplo (com *logs* de compilação omitidos). Você pode utilizar esse arquivo para validar os resultados da sua própria aplicação.

Para agilizar testes iniciais, recomendamos que sua aplicação seja configurada com um limite reduzido de resultados.

Entrega e publicação

Para entregar o código desenvolvido você deve seguir as instruções fornecidas durante o seu processo seletivo. Caso você queira disponibilizar seu código publicamente, solicitamos gentilmente que quaisquer referências à Axur sejam removidas de todos os arquivos antes da publicação.