

Desafio Técnico Luiza Labs

Bem vindo ao desafio técnico do Luiza Labs e muito obrigado por se interessar em fazer parte do Magazine Luiza.

O objetivo deste desafio é avaliar seu conhecimento sobre algoritmos de indexação e de busca em massas de dados, avaliar a qualidade do seu código em termos de escrita, padronização e performance na utilização de recursos e estruturas de dados.

Introdução

Neste desafio você precisará fazer download de um conjunto de arquivos que possuem dados de filmes, sendo eles o título, ano, elenco e equipe que pertencem ao filme. Este conjunto de arquivos possui, cada um, informações do filme que o intitula. Os dados estão no arquivo `movies.zip` enviado.

Exemplo:

Nome do arquivo	Conteúdo do arquivo
<code>donald-s-dream-voice.txt</code>	<code>walt disney duck donald</code>

Seu desafio é criar um programa que busque por uma sentença em todos estes arquivos e exiba quantos e quais arquivos possuem esta palavra chave. A ordenação dos arquivos deve ser feita em ordem crescente e alfabética (“a” para “z”). O seu programa deve ser executado a partir da linha de comando do terminal (Linux/MS-Windows/MacOS/Unix) e deve permitir que a sentença que deverá ser encontrada seja passada como parâmetro de execução.

Exemplo de comando de execução com a sentença "walt disney" sendo passada para o programa:

Linguagem	Exemplo
Python	<code>python search.py "walt disney"</code>
Java	<code>java -jar search.jar "walt disney"</code>
Javascript	<code>node search.js "walt disney"</code>

Exemplo de resposta do programa:

Foram encontradas 53 ocorrências pelo termo "walt disney".

Os arquivos que possuem "walt disney" são:

data/a-cowboy-needs-a-horse.txt
data/alice-and-the-three-bears.txt
data/alice-helps-the-romance.txt
data/alice-s-fishy-story.txt
data/alpine-climbers.txt
data/billposters.txt
data/bone-trouble.txt
data/bootle-beetle.txt
data/canine-caddy.txt
data/clown-of-the-jungle.txt
data/cock-o-the-walk.txt
data/cured-duck.txt
data/dog-watch.txt
data/donald-s-dream-voice.txt
data/donald-s-snow-fight.txt
data/dude-duck.txt
data/el-gaicho-goofy.txt
data/fall-out-fall-in.txt
data/father-noah-s-ark.txt
data/fathers-are-people.txt
data/football-now-and-then.txt
data/funny-little-bunnies.txt
data/gallopin-gaicho.txt
data/good-scouts.txt
data/goofy-and-wilbur.txt
data/great-guns.txt
data/how-to-be-a-sailor.txt
data/how-to-have-an-accident-at-work.txt
data/jiminy-cricket-s-christmas.txt
data/mickey-s-elephant.txt
data/moochie-of-pop-warner-football.txt
data/on-ice.txt
data/pantry-pirate.txt
data/perri.txt
data/self-control.txt
data/sky-scrappers.txt
data/society-dog-show.txt
data/spare-the-rod.txt
data/tall-timber.txt
data/the-four-musicians-of-bremen.txt
data/the-hockey-champ.txt
data/the-jazz-fool.txt
data/the-little-house.txt
data/the-mail-pilot.txt
data/the-merry-dwarfs.txt

```
data/the-nifty-nineties.txt
data/the-pet-store.txt
data/tommy-tucker-s-tooth.txt
data/tugboat-mickey.txt
data/two-gun-mickey.txt
data/undiscovered-walt-disney-world.txt
data/wide-open-spaces.txt
data/working-for-peanuts.txt
```

Requisitos

Você deve ter o código coberto por testes unitários e deve utilizar somente ferramentas nativas da linguagem que você escolher no desenvolvimento do algoritmo. É desejável que o código esteja versionado, preferencialmente usando GIT.

Para os testes, você pode utilizar qualquer framework, desde que este crie os relatórios para nossa análise.

Seu projeto deve ser entregue com um arquivo README.md que contenha as seguintes informações básicas:

- Dependências
- Versões
- Instruções de instalação
- Execução de testes
- Execução do programa.

Qualquer informação que você achar necessário colocar neste documento, será bem-vinda.

Sobre os requisitos funcionais, lembre-se que as palavras digitadas em um sistema de busca não necessariamente estão na mesma ordem em que elas estão salvas nos arquivos. O "match" das palavras deve utilizar o critério "AND", ou seja, ambas devem existir no arquivo, independente da ordem em que se encontram.

Exemplo:

Nome do arquivo	Conteúdo do arquivo	Termos buscados	Está no resultado?
donald-s-dream-voice.txt	walt disney duck donald	walt disney	Sim
donald-s-dream-voice.txt	walt disney duck donald	walt donald	Sim
donald-s-dream-voice.txt	walt disney duck donald	donald	Sim
donald-s-dream-voice.txt	walt disney duck donald	waltdisney	Não
donald-s-dream-voice.txt	walt disney duck donald	walt mickey	Não

Sobre os requisitos não funcionais, as seguintes informações devem ser seguidas:

Requisitos	Descrição
Desempenho	Depois das informações carregadas, o tempo máximo de resposta de uma busca é de 0,5 segundos (ou 500 milisegundos).
Manutenção	Código de fácil entendimento e desacoplado para ser o mais adaptativo possível (melhoria contínua).

Observação: Você pode pré-processar e preparar as informações dos arquivos em um outro sistema, script ou função, desde que ele venha incluído em seu GIT e tenha instruções de uso.