


| | |
|--|---|
|  INSTITUTO FEDERAL Brasília | Instituto Federal de Brasília Campus Taguatinga Superior em Computação |
| | Atividade 2 - Paralegrep Sistemas Operacionais - 2024/2 Professor João Victor de A. Oliveira |

Data de entrega: 20/12/2024

Valor: 30% da nota final da disciplina

Quantidade de membros por grupo: 3

Linguagem de Programação: C padrão ou C++

Sistema Operacional padrão para a correção: Ubuntu

Biblioteca de paralelismo a ser usada: pthreads.h


Formato do arquivo de entrega

O arquivo deve estar em formato .zip ou .7z com o nome formado pela concatenação de “T1_SO_”) e os primeiros nome de cada integrante. Ex.: T1_SO_joaopedromaria.zip. Além disso, deve conter um documento de texto denominado README.txt que deve conter o nome completo de cada membro do grupo, uma descrição de como compilar e executar o programa e uma descrição sucinta de como o código está organizado e de como ele funciona.

Paralegrep

O comando grep é usado em ambientes UNIX para encontrar padrões em arquivos. Por exemplo:

grep “sistemas” arquivo1.txt

| | |
|--|---|
|  INSTITUTO FEDERAL Brasília | Instituto Federal de Brasília Campus Taguatinga Superior em Computação |
| | Atividade 2 - Paralegrep Sistemas Operacionais - 2024/2 Professor João Victor de A. Oliveira |

retorna todas as ocorrências do termo “sistemas” em linhas do arquivo1.txt. Este comando se torna bastante útil quando temos uma coleção de arquivos de texto em uma pasta e precisamos encontrar alguma informação a partir de uma dada ocorrência de um determinado termo. Por exemplo:

```
grep “sistemas” *.txt
```


retorna todas as ocorrências de “sistemas” em linhas de todos os arquivos .txt da pasta corrente. Caso esteja fazendo uma pesquisa de quantas ocorrências de um termo ocorre em um dado arquivo (ou conjuntos de arquivos), basta adicionar o argumento -c (count) no grep:

```
grep “sistemas” *.txt -c
```

O primeiro trabalho de Sistemas Operacionais tem como objetivo criar um pequeno sistema multithread de busca de padrões em um conjunto de arquivos. Este sistema se chama **paralegrep** e deve ser implementado em linguagem C.

Especificações do sistema paralegrep

O sistema paralegrep recebe como argumento de entrada uma palavra de interesse (sem espaços e apenas letras). Exemplo de chamada do sistema paralegrep via terminal:

| | |
|--|---|
|  | Instituto Federal de Brasília Campus Taguatinga Superior em Computação |
| | Atividade 2 - Paralegrop Sistemas Operacionais - 2024/2 Professor João Victor de A. Oliveira |

./paralegrop sistemas

Ao iniciar o sistema paralegrop, este deve verificar quantas ocorrências do termo “sistemas” (ou outro termo definido pelo usuário) aparece em cada arquivo do diretório “fileset” que se encontra dentro do diretório corrente (que possui o programa ./paralegrop).

Além disso, deve-se imprimir na saída padrão (stdout) um ranking mostrando o top 10 arquivos com mais ocorrências do termo definido pelo usuário.


O programa deve permanecer em execução, pois, caso algum arquivo seja alterado ou incluído no diretório fileset, o ranking de quantidade de termos encontrados em cada arquivo deve ser atualizado.

Para fazer este sistema, deve-se usar 3 tipos de threads:

- Thread despachante;
- Threads operárias;
- Thread ranking.

A seguir são descritos as características e funcionalidades de cada tipo de thread.


Thread Despachante

| | |
|--|--|
|  | <p>Instituto Federal de Brasília Campus Taguatinga Superior em Computação</p> |
| | <p>Atividade 2 - Paralegrip Sistemas Operacionais - 2024/2 Professor João Victor de A. Oliveira</p> |

- Deve monitorar o diretório “*fileset*”, que deve estar armazenado dentro do diretório corrente;
 - A cada 5 segundos o thread despachante deve verificar se houve alterações ou inclusões de arquivos no diretório;
- Caso haja novos arquivos ou arquivos alterados, deve, para cada um desses arquivos, “acordar” alguma de suas 10 threads operárias e pedir a ela que conte quantas ocorrências do termo definido pelo usuário existem nestes arquivos;
 - Caso todas as threads operárias estejam trabalhando, deve-se esperar até que uma thread operária esteja disponível.

Threads Operárias

- Recebem um arquivo da thread despachante e devem contar quantas ocorrências do termo definido pelo usuário existem neste arquivo;
- O resultado deve ser enviado junto com o nome de arquivo à thread ranking (por exemplo, pode-se definir uma estrutura de dados que recebe este dado, sendo que esta estrutura de dados é monitorada pela thread ranking);

| | |
|--|---|
|  INSTITUTO FEDERAL Brasília | Instituto Federal de Brasília Campus Taguatinga Superior em Computação |
| | Atividade 2 - Paralegrip Sistemas Operacionais - 2024/2 Professor João Victor de A. Oliveira |

Thread Ranking

- Recebem das threads operárias a quantidade de ocorrências do termo definido pelo usuário;
- Após receber um novo dado, esta deve atualizar o ranking e imprimir em tela o atual top 10 de arquivos com mais ocorrências de um dado termo.

Observação

- Como toda especificação de um software, dúvidas podem surgir e estas precisam ser esclarecidas diretamente com os stakeholders (no caso o professor). Faz parte da avaliação a busca de informações adicionais para o pleno cumprimento e funcionamento do software demandado.