

Ferramenta de Busca em Texto

Rodrigo Lopes de Carvalho

Identificação

Rodrigo Lopes de Carvalho - único integrante.
Implementou os algoritmos e a interface da cli.

Implementação

Algoritmos Implementados:

- KMP
- Sellers
- Aho-Corasick
- Força Bruta

O algoritmo KMP será utilizado quando só é informado UM padrão.
Para mais padrões, o algoritmo de Aho-Corasick é utilizado, pois ele suporta tempo $O(n)$ com vários padrões ao mesmo tempo.

O algoritmo Sellers é utilizado quando for informada uma distância de edição.

O algoritmo de força bruta só será utilizado caso o usuário use a flag "-a" para forçar um algoritmo.

Bugs:

A implementação do algoritmo Wu-Manber não foi bem sucedida. O código está disponível como registro histórico.

Testes

Os tempos de execução dos testes foram medidos usando a ferramenta “time”.

O arquivo de entrada utilizado foi "english.50MB" disponível em <http://pizzachili.dcc.uchile.cl/texts/nlang/>

Outro arquivo, `english200`, é a repetição do mesmo texto, 4 vezes.

KMP:

```
bin/pmt -a k inheritance ~/Downloads/english.50MB 2.49s user 0.07s system 91% cpu 2.785 total
bin/pmt -a k inheritance ~/Downloads/english.50MB 2.41s user 0.07s system 90% cpu 2.722 total
bin/pmt -a k preposterous ~/Downloads/english.50MB 2.27s user 0.04s system 95% cpu 2.409 total
bin/pmt -a k preposterous ~/Downloads/english200 9.38s user 0.18s system 95% cpu 9.982 total
```

Sellers:

```
bin/pmt -a s inheritance ~/Downloads/english.50MB 11.87s user 0.65s system 95% cpu 13.164 total
bin/pmt -a s inheritance ~/Downloads/english.50MB 12.22s user 0.85s system 94% cpu 13.852 total
bin/pmt -a s preposterous ~/Downloads/english.50MB 12.32s user 0.69s system 97% cpu 13.281 total
bin/pmt -a s preposterous ~/Downloads/english200 52.26s user 4.20s system 93% cpu 1:00.63 total
```

Aho-Corasick:

```
bin/pmt -a a inheritance ~/Downloads/english.50MB 5.34s user 0.08s system 94% cpu 5.738 total
bin/pmt -a a inheritance ~/Downloads/english.50MB 5.31s user 0.08s system 94% cpu 5.718 total
bin/pmt -a a preposterous ~/Downloads/english.50MB 4.96s user 0.06s system 97% cpu 5.162 total
bin/pmt -a a preposterous ~/Downloads/english200 19.26s user 0.17s system 97% cpu 19.949 total
```

Força Bruta:

```
bin/pmt -a d inheritance ~/Downloads/english.50MB 2.41s user 0.07s system 92% cpu 2.688 total
bin/pmt -a d inheritance ~/Downloads/english.50MB 2.44s user 0.07s system 91% cpu 2.744 total
bin/pmt -a d preposterous ~/Downloads/english.50MB 2.28s user 0.04s system 95% cpu 2.419 total
bin/pmt -a d preposterous ~/Downloads/english200 9.31s user 0.14s system 96% cpu 9.745 total
```

Conclusão

Como se pode ver nos resultados, os algoritmos KMP e Força Bruta tiveram resultados similares.

A implementação do algoritmo Sellers, porém, teve um desempenho muito abaixo do previsto.