

EP5 - Ocorrência de palavras

Entrega: 11/12/2006

Seu EP5 deve receber na linha de comando o nome de um arquivo texto, por exemplo, `texto.txt`. A partir daí, ele espera que o usuário forneça palavras no `stdin`, precedidas de `+` ou `-`, uma palavra por linha. (i) Se uma palavra é precedida de `+`, seu programa deve imprimir todas as linhas de `texto.txt` que contêm a palavra. (ii) Se uma palavra é precedida de `-`, seu programa deve imprimir todas as linhas de `texto.txt` que *não* contêm a palavra. Quando o usuário estiver satisfeito, ele fornecerá uma linha que começa com `.` para pedir o fim da execução de seu programa.

Exemplo. Suponha que `texto.txt` seja

```
When my love swears that she is made of truth
I do believe her, though I know she lies,
That she might think me some untutor'd youth,
Unlearned in the world's false subtleties.
Thus vainly thinking that she thinks me young.
Although she knows my days are past the best,
Simply I credit her false speaking tongue:
On both side thus is simple truth supress'd:
And wherefore says she not she is unjust?
And wherefore says not I that I am old?
O! love's best habit is in seeming trust,
And age in love loves not to have years told:
Therefore I lie with her and she with me,
And in our faults by lies we flatter'd be.
```

Esta poderia ser uma execução de seu programa:

```
yoshi@erdos:~/Main/www/2006ii/mac122a/exx/sink$ ep5 texto.txt
+the
Unlearned in the world's false subtleties.
Although she knows my days are past the best,
+love
When my love swears that she is made of truth
O! love's best habit is in seeming trust,
```

```
And age in love loves not to have years told:
-she
Unlearned in the world's false subtleties.
Simply I credit her false speaking tongue:
On both side thus is simple truth supress'd:
And wherefore says not I that I am old?
O! love's best habit is in seeming trust,
And age in love loves not to have years told:
And in our faults by lies we flatter'd be.
.
yoshi@erdos:~/Main/www/2006ii/mac122a/exx/sink$
```

Implementação. Seu programa deve necessariamente implementar uma tabela de símbolos (para as palavras que ocorrem no texto). Seu programa deve ser capaz de processar arquivos `texto.txt` bastante grandes (como livros inteiros). Para cada palavra dada no `stdin` (`+the`, `-she`, etc), a saída deve ser produzida “rapidamente”.

Dois versões. Você pode implementar duas versões de seu EP5. A primeira, digamos `ep5.a.c`, deve implementar a tabela de símbolos com ABBs. A sua segunda implementação, digamos `ep5.b.c`, *que é opcional e valerá um bônus*, deve implementar a tabela de símbolos com tabelas de espalhamento (*hashing*). Se você fizer as duas versões, não deixe de comparar a eficiência das duas versões, e escreva um pequeno relatório com suas conclusões.