

Processamento de Linguagens e Compiladores 3º ano
Expressões Regulares e GAWK
Grupo 14

Artur Queiroz
(A77136)

Rafael Fernandes
A78242

Rafaela Pinho
A77293

15 de Outubro de 2017

Resumo

Este trabalho foca os conceitos básicos do funcionamento GAWK e das expressões regulares utilizadas para descrever padrões. Neste relatório descrevemos as decisões tomadas e as dificuldades encontradas, bem como pequenos exemplos para que qualquer um que o leia perceba facilmente como funciona o nosso projeto.

Conteúdo

1	Introdução	2
2	Ficheiros de internacionalização	3
2.1	Descrição informal do problema	3
2.2	Especificação do Requisitos	3
3	Concepção/desenho da Resolução	4
3.1	Estruturas de Dados	4
3.2	Algoritmos	4
4	Codificação e Testes	5
4.1	Problemas de implementação, Decisões e Alternativas	5
4.1.1	Problemas de implementação	5
4.1.2	Decisões	5
4.1.3	Alternativas	5
4.2	Testes realizados e Resultados	5
5	Conclusão	6
6	Código do Programa	7

Capítulo 1

Introdução

Neste trabalho vamos filtrar vários ficheiros de internacionalização, utilizando expressões regulares e o sistema de filtragem de texto GAWK. Uma parte importante do sistema linux são as expressões regulares e a capacidade de procurar num input um determinado padrão, o que facilita muitas das operações. As expressões regulares são uma sequência de caracteres, sucinta e flexível, que identifica palavras ou padrões de caracteres. Combinando-as com o sistema de filtragem, conseguimos extrair informações dos ficheiros POs que são documentos que contêm traduções de uma língua para outra. Este trabalho tem como objetivos aprofundar o conhecimento do GAWK, do sistema Linux e as suas ferramentas, bem como aprender melhor o funcionamento das expressões regulares.

Capítulo 2

Ficheiros de internacionalização

2.1 Descrição informal do problema

É necessária um "script" que permita:

- a) Analisar vários ficheiros de formato PO, e devolver o número de traduções e os seus tradutores, bem como alguns metadados acharmos relevantes.
- b) Através de ficheiros PO's criar dicionários de triangulação de Português para Francês.
- c) Reformatar ficheiros de formato PO pra LaTeX.

2.2 Especificação do Requisitos

Os requisitos mínimos deste trabalho são saber utilizar Linux, bash e GAWK.

Capítulo 3

Concepção/desenho da Resolução

3.1 Estruturas de Dados

3.2 Algoritmos

Utilizamos as seguintes expressões regulares:

- i) `/^msgid/`
- ii) `/^"Language:/`
- iii) `<.*@.*> / && !/^"Report/ && !/FIRST/ && !/^"Language/`
- iv) `/^"Language-Team: *\n"/`
- v) `/^"Language-Team:/`
- vi) `/portugu[êe]se?s?/`
- vii) `/^msgid[\t]*/`
- viii) `/^msgstr[\t]*/`
- ix) `/^msgid */`
- x) `/^msgstr */`

Capítulo 4

Codificação e Testes

4.1 Problemas de implementação, Decisões e Alternativas

4.1.1 Problemas de implementação

De um modo geral no trabalho, tivemos problemas com a formatação de alguns ficheiros, dado não se encontrarem na formatação prevista, e com o LaTeX.

4.1.2 Decisões

No problema 1 proposto no exercício, decidimos ignorar os "msgid" que não contivessem nenhuma mensagem ou que contivessem mais de uma tradução. Escolhemos como metadados mais importantes a linguagem, a percentagem de blocos, a quantidade de blocos por ficheiro, e quantos e quem são os tradutores. Decidimos também que a linguagem só está no "Language". No problema 2 decidimos utilizar a "Language-team" para definir a linguagem de tradução, pois nenhum dos ficheiros apresentava linguagem no campo "Language". Usamos uma matriz descritiva, sendo as linhas a frase em inglês e os descritivos das colunas a linguagem portuguesa ou francesa. Assumimos só a linguagem portuguesa e a francesa e imprimimos unicamente o que tem tradução em triangulação. No último problema ignoramos o "

n" e o "#", devido ao LaTeX, substituindo por e por -, respetivamente.

4.1.3 Alternativas

Uma alternativa seria modificar os ficheiros manualmente de modo a estarem no formato certo. Outra alternativa seria fazer inúmeros casos especiais.

4.2 Testes realizados e Resultados

Mostram-se a seguir alguns testes feitos (valores introduzidos) e os respectivos resultados obtidos:

Capítulo 5

Conclusão

Com este trabalho concluímos que as expressões regulares são excelentes auxiliares na procura do que pretendemos em vários ficheiros simultaneamente. Para além disso, este trabalho permitiu a familiarização com o LaTeX, já que até então não era uma ferramenta utilizada regularmente. Como perspetiva futura, e mencionado anteriormente, sugerimos a alteração dos ficheiros que não estão com o formato correto, de modo a não conterem "lixo" no resultado da nossa procura. Esta opção seria a mais sensata, dado que criar casos especiais para esses ficheiros é um procedimento mais complexo.

Capítulo 6

Código do Programa

Script do exercício 2.5 a)

Script do exercício 2.5 b)

Script do exercício 2.5 c)