

Processamento de Linguagens e Compiladores
Processador de Róis de Confessados
Relatório de Desenvolvimento

André Sá (76361)

João Bastos (47419)

Pedro Sá (78164)

18 de Novembro de 2018

Resumo

O trabalho apresentado consiste na criação de um processador de texto que a partir de Róis de Confessados disponibilizados calcula a frequência de alguns padrões.

Conteúdo

1	Introdução	2
1.1	Enquadramento	2
1.2	Estrutura do Relatório	2
2	Análise e Especificação	3
2.1	Descrição informal do problema	3
2.2	Especificação dos Requisitos	3
3	Concepção/desenho da Resolução	4
3.1	Estruturas de Dados	4
3.2	Algoritmos	4
4	Testes	5
4.1	Alternativas, Decisões e Problemas de Implementação	5
4.2	Testes realizados e Resultados	5
5	Conclusão	8
A	Código do Programa	9

Capítulo 1

Introdução

1.1 Enquadramento

Os róis de confessados eram listas nominativas elaboradas durante a quaresma a fim de constatar quais fiéis se confessaram e comungaram no último ano. Este texto determinava que nos róis constassem os nomes de todos os paroquianos, estado matrimonial, se eram crismados, se eram menores de idade, entre outras informações consideradas relevantes. Os róis de confessados deveriam listar os fregueses de cada domicílio, separando as informações com uma linha. O processador texto que iremos criar será implementado sobre um exemplo de rol de confessados.

1.2 Estrutura do Relatório

Neste relatório começamos com a Introdução (Capítulo 1) onde, de uma forma breve tentamos descrever em

Explicar como está organizado o documento, referindo os capítulos existentes em [?] e a sua articulação explicando o conteúdo de cada um. No capítulo ?? faz-se uma análise detalhada do problema proposto de modo a poder-se especificar as entradas, resultados e formas de transformação.

etc. ...

No capítulo 5 termina-se o relatório com uma síntese do que foi dito, as conclusões e o trabalho futuro

Capítulo 2

Análise e Especificação

2.1 Descrição informal do problema

Foi-nos proposto construir um ficheiro Gawk, para processar um rol de confessados, com o intuito de calcular o número de certas ocorrências.

2.2 Especificação dos Requisitos

O Rol de Confessados que nos foi facultado tem um registo por linha. Cada campo do registo está separado por ":", sendo que 6 têm o seguinte conteúdo:

1º campo (\$1) Número do registo;

2º campo (\$2) Data do registo;

3º campo (\$3) Nome do paroquiano a registar;

4º campo (\$4) Pai do paroquiano;

5º campo (\$5) Mãe do paroquiano;

6º campo (\$6) Informação complementar.

e o último campo apresenta-se sempre vazio.

Todos os registos têm os primeiros 3 campos preenchidos, podendo os restantes estar omissos. Apresentemos o exemplo de um registo completo para facilitar na compreensão do exposto acima e nas seguintes referências a registos do Rol de Confessados.

Foi-nos então proposto sobre este texto fazermos o seguinte processamento:

1. Calcular a frequência de processos por ano. Dentro do 2ª campo
2. Calcular a frequência de nomes.
3. Calcular a frequência dos vários tipo de relação.

Capítulo 3

Concepção/desenho da Resolução

3.1 Estruturas de Dados

Um array associativo é uma estrutura de dados - análoga aos dicionários - composta por um conjunto não ordenado de itens formados por um par chave e valor, em que cada chave possui um valor associado.

O Gawk permite-nos utilizar strings como índices. Assim, optamos pela utilização de arrays associativos para contar a ocorrência de determinadas strings sendo as strings as chaves e o valor correspondente ao número de ocorrências de cada uma.

3.2 Algoritmos

Visto não haver vontade ou qualquer necessidade de escrever merda nenhuma aqui, optamos simplesmente por nada escrever. 20 valores!!

Capítulo 4

Testes

4.1 Testes realizados e Resultados

Devido ao facto do rol de confessados ter bastante informação para ser processada (≈ 37000 linhas), o grupo decidiu para efeitos de apresentação de testes e resultados desse processamento, apresentar neste relatório uma porção do rol de confessados representativa de todo o tipo de variações de apresentação de tal informação.

Ao processar o rol de confessados verificamos que existe algumas variações de apresentação da informação, isto é, existe registos (linhas do rol de confessados) que os campos não estavam todos preenchidos, mais especificamente o campo 6, e por vezes como foi referido na secção 2.2, esse campo contém informação que varia entre parentesco e/ou uma descrição relevante para aquele registo.

Abaixo encontra-se a tal porção do rol de processos, onde cada linha representa um registo desse rol e contém os campos referidos na secção 2.2.

1	616::1778-01-14::Pedro Jose Sousa::Lourenco Rodrigues::Brisida Sousa:::
2	598::1722-07-29::Pedro Jose Sousa::Pedro Sousa Falcao::Mariana Sousa Pereira::Manuel Felix Sousa, Irmão. Proc.8224.::
3	606::1733-02-09::Pedro Jose Teixeira::Jeronimo Araujo Pinheiro::Francisca Teixeira::Doc. danificado.::
4	384::1732-12-02::Pedro Jose Teixeira Azevedo::Cristovao Azevedo::Maria Teixeira:::
5	617::1783-02-12::Pedro Jose Vieira::Manuel Vieira Sousa::Teresa Maria Pimenta::Jose Vieira Sousa, Irmão. Proc.24811.::
6	599::1722-08-13::Pedro Lama::Joao Lama::Ana Gomes:::
7	589::1691-05-16::Pedro Leite Ferreira::Antonio Ferreira::Maria Leite::Em Anexo: Inquiricao feita em 1694/08/11.::
8	591::1703-06-20::Pedro Lenoble::Pedro Lenoble::Joana Noble::Residente em Braga. Seus pais sao naturais de Londres.::
9	587::1681-02-12::Pedro Lima Antunes::Francisco Sampaio Antunes::Justa Goncalves
10	1246::1695-05-07::Bento Silva Teles::Henrique Sousa Teles::Maria Silva::Fez Inquiricao para Conego da Se de Braga, em 1708/07/24. Ver TomoV, fls.430. Antonio Silva Teles, Joao Silva Teles, Manuel Sousa Teles e Matias Ribeiro, Irmaos. Proc.7984. Francisco Silva Teles, Sobrinho Paterno. Proc.28871.::
11	85::1689-08-27::Jose Soares Pereira::Joao Soares Pereira::Margarida Silva Vasconcelos::Gaspar Silva Pereira, Luis Soares Bacelar e Pedro Beca Pereira, Irmaos. Proc.27470.::
12	578::1901-10-24::Antonio Pereira Lomba::Manuel Pereira Lomba::Rosa Maria Alves::Antonio Pereira Lomba, Tio Paterno. Proc.13040.::
13	561::1845-04-29::Antonio Pereira Silva Junior::Antonio Pereira Silva::Maria Jose Alvares::Jose Joaquim Silva Pereira, Irmão. Proc.26187. Manuel Lopes Oliveira, Tio Materno. Proc.10442.::

De seguida temos o ficheiro com o resultado da frequência de processos por ano:

1	Frequencia	de processos por ano:
2	1681	1
3	1689	1
4	1691	1
5	1695	1
6	1703	1
7	1722	2
8	1732	1
9	1733	1
10	1778	1
11	1783	1
12	1845	1
13	1901	1

De seguida temos o ficheiro com o resultado da frequência de nomes:

1	Frequencia	de nomes:
2	Lomba	2
3	Soares	2
4	Lama	2
5	Ana	1
6	Gomes	1
7	Noble	1
8	Brisida	1
9	Vasconcelos	1
10	Cristovao	1
11	Rodrigues	1
12	Antonio	4
13	Margarida	1
14	Joao	2
15	Pinheiro	1
16	Falcao	1
17	Azevedo	2
18	Henrique	1
19	Pimenta	1
20	Pedro	11
21	Francisca	1
22	Junior	1
23	Antunes	2
24	Teixeira	4
25	Lenoble	2
26	Francisco	1
27	Sampaio	1
28	Joana	1
29	Jeronimo	1
30	Alvares	1
31	Rosa	1
32	Manuel	2
33	Araujo	1
34	Ferreira	2
35	Sousa	7
36	Maria	6
37	Alves	1
38	Jose	7
39	Pereira	7
40	Mariana	1
41	Justa	1
42	Bento	1

43	Teles	2	
44	Lima	1	
45	Lourenco		1
46	Teresa	1	
47	Goncalves		1
48	Vieira	2	
49	Silva	5	
50	Leite	2	

De seguida temos o ficheiro com o resultado da frequência de vários tipos de relação:

1	Frequencia dos varios tipos de relacao:	
2	sobrinho	1
3	tio	2
4	irmao	5

Capítulo 5

Conclusão

Este projecto mostrou-nos que o Gawk permite-nos programar filtros de texto e relatórios sobre informação de forma muito acessível, principalmente quando existe muito informação para ser processada.

Este projeto está finalizado para o que nos foi proposto. E se no futuro houver necessidade de haver uma maior filtragem simplesmente é seguir a estrutura dos filtros presentes e adaptá-los, o que é fácil tendo em conta que o código é bastante simples em termos de funcionamento e bastante legível.

Apêndice A

Código do Programa

Lista-se a seguir o código Awk do programa que foi desenvolvido.

```
BEGIN {
    FS="::";
    IGNORECASE=1
}

/^\s*$/ { next; }

{ # Alinea (a)
    split($2, data, /-/);
    alA[data[1]]++;
}

{ # Alinea (b)
    for (i = 3; i < 6; i++) {
        # Tirar Expostos/Solteiras/etc
        gsub(/,.*$/, "", $i);

        split($i, nomes, /\s+/);

        for (nome in nomes)
            alB[nomes[nome]]++;
    }
}

# Alinea (c)
$6 ~ /\,<avo>([ ]+[mp]atern[oa]s)?\./ { alC["avo"]++; }
$6 ~ /\,<irmaos>([ ]+[mp]atern[oa]s)?\./ { alC["irmao"]++; }
$6 ~ /\,<irmas>([ ]+[mp]atern[oa]s)?\./ { alC["irma"]++; }
$6 ~ /\,<mae>\./ { alC["mae"]++; }
$6 ~ /\,<madrastas>([ ]+[mp]atern[oa]s)?\./ { alC["madrasta"]++; }
$6 ~ /\,<padrastos>([ ]+[mp]atern[oa]s)?\./ { alC["padrasto"]++; }
$6 ~ /\,<madrinhas>([ ]+[mp]atern[oa]s)?\./ { alC["madrinha"]++; }
$6 ~ /\,<padrinhos>([ ]+[mp]atern[oa]s)?\./ { alC["padrinho"]++; }
$6 ~ /\,<pai>\./ { alC["pai"]++; }
$6 ~ /\,<primas>([ ]+[mp]atern[oa]s)?\./ { alC["prima"]++; }
$6 ~ /\,<primos>([ ]+[mp]atern[oa]s)?\./ { alC["primo"]++; }
$6 ~ /\,<sobrinhas>([ ]+[mp]atern[oa]s)?\./ { alC["sobrinha"]++; }
```

```

$6 ~ /\,<sobrinhos?\>([ ]+[mp]atern[oa]s?)?\./ { alC["sobrinho"]++; }
$6 ~ /\,<tias?\>([ ]+[mp]atern[oa]s?)?\./ { alC["tia"]++; }
$6 ~ /\,<tios?\>([ ]+[mp]atern[oa]s?)?\./ { alC["tio"]++; }

function alineaA (outf)
{
    printf "Frequencia de processos por ano:\n" > outf;
    for (ano in alA)
        printf("%s\t%s\n", ano, alA[ano]) >> outf;
}

function alineaB (outf)
{
    printf "Frequencia de nomes:\n" > outf;
    for (nome in alB)
        printf("%s\t%s\n", nome, alB[nome]) >> outf;
}

function alineaC (outf)
{
    printf "Frequencia dos varios tipos de relacao:\n" > outf;
    for (parentesco in alC)
        printf("%s\t%s\n", parentesco, alC[parentesco]) >> outf;
}

END {
    alineaA("anos.txt");
    alineaB("nomes.txt");
    alineaC("parentescos.txt");
}

```