Processamento de Linguagens (3º ano de Mestrado Integrado em Engenharia Informática)

Trabalho Prático nº1 - Parte A

Relatório de Desenvolvimento

Daniel Malhadas (A72293) Alexandre Silva (A72502)

13 de Março de 2017

Resumo

O presente documento descreve um estudo de desenvolvimento de filtros de texto em **GAWK**. Como caso prático de estudo foi escolhido como fonte a filtrar um conjunto de dados relativos a transações no contexto da tecnologia **Via-Verde**. A fonte a filtrar é então um ficheiro de texto no formato **XML** com várias transações e, os filtros idealizados, consistem na organização de uma grande quantidade de informação vinda do referido ficheiro de forma a que o olho humano facilmente encontre pormenores do seu interesse, sendo assim apresentada a informação de forma clara, limpa e sucinta.

Área: Processamento de Linguagens, GAWK, XML

Conteúdo

1	Introdução	2
2	Análise e Especificação	4
	2.1 Descrição informal do problema	4
	2.2 Especificação dos Requisitos	4
	2.2.1 Dados	4
	2.2.2 Pedidos	5
3	Concepção/desenho da Resolução	7
	3.1 Estruturas de Dados	7
	3.2 Algoritmos	9
4	Codificação e Testes	10
	4.1 Alternativas, Decisões e Problemas de Implementação	10
	4.2 Testes realizados e Resultados	
5	Conclusão e Trabalho Futuro	18
\mathbf{A}	Código do Programa	19

Introdução

Supervisor: Pedro Rangel Henriques

Supervisor: José João

No decorrer da Unidade Curricular **Processamento de Linguagens**, foi proposto aos autores, pelos docentes supervisores, um trabalho prático com o intuito de desenvolver os seguintes objetivos:

- Aumentar a experiência de uso do ambiente Linux e de algumas ferramentas de apoio à programação;
- Aumentar a capacidade de escrever Expressões Regulares (ER) para descriçãao de padrões de frases;
- Desenvolver, a partir de ERs, sistemática e automaticamente Processadores de Linguagens Regulares, que filtrem ou transformem textos;
- Utilizar o sistema de produção para filtragem de texto GAWK.

É importante referir que no decorrer deste relatório teve-se também em consideração que grande parte dos problemas com o software hoje em dia baseia-se na má comunicação entre os desenvolvedores e os seus empregadores tentou-se então que todos as imagens e explicações neste relatório fossem sucintas, diretas ao assunto e simples de perceber. Desta forma emula-se a estrutura de um relatório que poderia ser relativo a um trabalho no mundo de tabalho fora da redoma que é a Universidade e onde o nosso empregador, mesmo com poucos conhecimentos de processamento de linguagens seria, mesmo assim, capaz de ler, perceber e fazer uma crítica constructiva sobre a direcção do projeto. Isto possibilita evitar interpretações erradas do objetivo por parte dos programadores e também permite ao empregador compreender melhor se realmente os seus objectivos podem ser todos realizados da forma que idealizou ou se são demasiado optimistas.

Estrutura do Relatório

explicar como está organizado o documento, referindo os capítulos existentes em e a sua articulação explicando o conteúdo de cada um. No capítulo 2 faz-se uma análise detalhada do problema proposto Organizamos então o relatório nos seguintes capítulos :

- Capítulo 1 Introdução Aqui são explicados sucintamente os objetivos gerais que se pretende com o projecto. Não os os objetivos concretos relativos á tecnologia ViaVerde, mas sim as capacidades globais que é expectável que os autores, como desenvolvedores do projeto, melhorem/desenvolvam. Fala-se também um pouco nos cuidados burucráticos tidos no processo de escrita do relatório bem como a sua organização estrutura.
- Capítulo 2 Análise e Especificação Começa-se por fazer uma descrição informal do problema a resolver, isto consiste na formulação do projecto já tocando nos pontos específicos que são relativos á teconologia ViaVerde.
 O input usado é depois especificado assim como os requisitos mínimos propostos pelos docentes supervisores do projeto, assim como os requisitos extra que os autores desenvolvedores acharam relevante adicionar.

- Capítulo 3 Concepção/Desenho da Solução De seguida é explicada a forma de iteração pelas linhas do ficheiro, sendo explicada a estrutura usada para guardar os dados necessários à medida que se itera pelo ficheiro e explicado também o raciocínio geral por detrás das soluções.
- Capítulo 4 Codificação e Testes Agora é mostrado o output de cada programa pormenorizadamente e demonstrado o seu respetivo método de uso. Referem-se também um pouco as dificuldades encontradas
- Capítulo 5 Conclusão e Trabalho Futuro Por fim faz uma breve análise de todo o projeto num panorama geral e avalia-se a solução obtida, mencionando também melhorias futuras que se poderiam vir a fazer caso o projeto não terminasse ainda.

Análise e Especificação

2.1 Descrição informal do problema

De forma a praticar os objetivos referidos no capítulo 1 realizou-se um projecto onde, de forma a operar sobre um ficheiro **XML** com informações relativas a transacções no contexto da tecnologia ViaVerde, se desenvolveram filtros de texto em **GAWK** que permitem a apresentação de informação pertinente ao utilizador de forma fácil, limpa e simples.

2.2 Especificação dos Requisitos

2.2.1 Dados

No ficheiro \mathbf{XML} referido anteriormente temos um extracto onde é identificado o cliente ao qual se refere e, de seguida, todas as transacções do mesmo.

Mostra-se agora o exemplo de dados que adjetivam um extracto:

- <EXTRACTO id="011114056/08/2015">
- 2 <MES_EMISSAO>Ago-2015</MES_EMISSAO>
- 3 <CLIENTE id="514714936">...</CLIENTE>
- <IDENTIFICADOR id="28876820811">
- 5 <MATRICULA>00-LJ-11</MATRICULA>
- 6 <REF_PAGAMENTO>1234567</REF_PAGAMENTO>

O campo CLIENTE expande para as seguintes linhas:

- 1 <CLIENTE id="514714936">
- <NIF>987653210</NIF>
- 3 <NOME>PEDRO MANUEL RANGEL SANTOS HENRIQUES</NOME>
- 4 <MORADA>RUA XXX</MORADA>
- 5 <LOCALIDADE>BRAGA</LOCALIDADE>
- 6 <CODIGO_POSTAL>4715-012 BRAGA</CODIGO_POSTAL>
- 7 </CLIENTE>

De seguida temos ainda várias transacções relativas ao extrato. Cada uma destas transacções pode ser de vários tipos distintos. No ficheiro utilizado surgem apenas dois tipos: **Portagens** e **Parques de estacionamento**. Mostra-se agora o exemplo de uma portagem:

```
<TRANSACCAO>
```

- 2 <DATA_ENTRADA>26-07-2015</DATA_ENTRADA>
- <HORA_ENTRADA>11:33</HORA_ENTRADA>
- 4 <ENTRADA>Povoa N-S</ENTRADA>
- 5 <DATA_SAIDA>26-07-2015</DATA_SAIDA>
- 6 < HORA_SAIDA>11:42</ HORA_SAIDA>
- <SAIDA>Angeiras N-S</SAIDA>
- <IMPORTANCIA>2,00</IMPORTANCIA>
- 9 <VALOR_DESCONTO>0,00</VALOR_DESCONTO>
- O <TAXA_IVA>23</TAXA_IVA>
- 11 <OPERADOR>I. de Portugal (N1)</OPERADOR>
- 12 <TIPO>Portagens</TIPO>
- 13 <DATA_DEBITO>05-08-2015</DATA_DEBITO>
- 14 <CARTAO>6749036</CARTAO>
- 15 </TRANSACCAO>

Note-se que é referido o local do pórtico de saída e o de entrada (que iremos referir de agora para a frente como saída e entrada apenas). Vemos agora um exemplo de um parque de estacionamento:

```
<TRANSACCAO>
```

- 2 <DATA_ENTRADA>06-08-2015</DATA_ENTRADA>
- < <HORA_ENTRADA>15:57</HORA_ENTRADA>
- 4 <ENTRADA/>
- 5 <DATA_SAIDA>06-08-2015</DATA_SAIDA>
- 6 <HORA SAIDA>17:08</HORA SAIDA>
- 7 <SAIDA>PQ A Sa Carn.I</SAIDA>
- «IMPORTANCIA»

 3,75</importancia»

 </p>
- 9 <VALOR_DESCONTO>0,00</VALOR_DESCONTO>
- 10 <TAXA_IVA>23</TAXA_IVA>
- <OPERADOR>ANA Aeroportos de Portugal. SA (AP)
- 12 <TIPO>Parques de estacionamento</TIPO>
- 13 <DATA_DEBITO>14-08-2015</DATA_DEBITO>
- 14 <CARTAO>6749036</CARTAO>
- 15 </TRANSACCAO>

Aqui vemos que não temos o campo relativo à entrada. Sendo então o campo de sáida referente ao local do parque de estacionamento.

2.2.2 Pedidos

Certos requisitos mínimos foram propostos pelos docentes supervisores do projeto:

- A) Calcular o número de 'entradas' em cada dia do mês.
- B) A lista de locais de 'saída'.
- C) Calcular o total gasto no mês.
- D) Calcular o total gasto no mês apenas em 'parques'.

para além destes requisitos mínimos, desenvolveram-se também alguns extras. Extendeu-se os quatro requisitos iniciais e adicionou-se ainda um outro.

A) Calcular o número de 'entradas' em cada data distinta, calcular quantas em cada dia distinto, quantas em cada mês distinto e quantas em cada ano distinto.

- B) Determinar lista de locais de 'saída' e, para cada local de saída distinto, quantas saídas no mesmo se realizaram.
- C) Calcular o total gasto em cada data, em cada dia distinto, em cada mês distinto e em cada ano distinto.
- **D)** Calcular o total gasto em cada data, em cada dia distinto, em cada mês distinto e em cada ano distinto, mas apenas em 'parques'.
- E) Construir uma árvore de ficheiros HTML, ligados entre si através de links, que apresentam de forma apelativa toda a informação presente no ficheiro XML. Deve ainda indicar o tempo de duração de cada 'viagem', o tempo de estadia em cada 'parque', assim como o tempo total para ambos. As viagens devem incluir um link para o respetivo percurso no google maps e os parques um link para o respetivo local no google maps.

Concepção/desenho da Resolução

3.1 Estruturas de Dados

Todos os programas desenvolvidos neste projeto partilham uma estrutura semelhante. Inicialmente têm um bloco **BEGIN** onde são feitos prints iniciais de apoio ao utilizador e declaradas algumas variáveis importantes para todo o programa, como contadores e nomes de ficheiros.

Segue-se um exemplo de um bloco deste tipo usado no programa A:

De seguida têm um bloco intermédio que será o usado para iterar pelo ficheiro de input. Este bloco consiste na seguinte estrutura: **\EXPRESSAO REGULAR\{Operar sobre a linha com a expressão regular procurada}**. Com este bloco exemplificado as instruções no interior das chavetas serão aplicadas para cada linha que contenha a expressão regular dentro dos 'backlashes'.

Mostra-se agora um exemplo deste bloco, também tirado do programa A:

Por fim temos um bloco **END** onde fazemos os prints finais de apresentação de resultados ao utilizador. De novo mostramos um exemplo tirado do programa A:

```
print "\t| Entradas em cada data"
print "\t|----"
for (i in datas)
   print "\t| Na Data " i ": " datas[i] " entradas";
print "\t|----"
print "\t| Entradas em cada dia"
print "\t|----"
for (i in dia)
   print "\t| No dia " i ": " dia[i] " entradas";
print "\t|----"
print "\t| Entradas em cada mes"
print "\t|-----"
for (i in mes) {
   if(length(i)==0) print "\t| No mes null: " mes[i] " entradas";
            print "\t| No mes " i ": " mes[i] " entradas";
   else
print "\t|----"
print "\t| Entradas em cada ano"
print "\t|----"
for (i in ano) {
   if(length(i)==0) print "\t| No ano null: " ano[i] " entradas";
            print "\t| No ano " i ": " ano[i] " entradas";
   }
print "\t|----"
print "\t| Total Entradas: " entradas
print "\t\\_____/"
```

3.2 Algoritmos

Podemos observar na secção anterior que no interior do bloco intermédio do programa A se utiliza a função split para operar sobre a linha. Este método de operar sobre a linha repete-se da mesma forma nos restantes programas, segue-se então uma explicação do uso da função split:

Figura 3.1: Funcionamento da Função split em GAWK

Esta função irá guradar o resultado no segundo argumento e, este resultado será um array de strings com a string resultado (primeiro argumento) dividida pelo delimitador que é o terceiro argumento.

Nos outros programas surge por vezes também a necessidade de operar na linha abaixo da atual. Para isso temos mais uma função que podemos usar chamada **getline** que coloca a linha diretamente abaixo na variável \$0. Em certas alturas esta função é o ideal, mas noutras alturas em que a linha está muito em baixo em relação á nossa posição atual ou então quando não sabemos quantas linhas ainda faltam até à linha que pretendemos encontrar, usamos a função **index** em conjunção com **getline**. Por exemplo:

Com este ciclo vamos sempre percorrendo linha a linha até a linha atual (\$0) tiver a string pretendida (DATA_DEBITO). Em situações em que não temos a certeza se uma linha realmente existe podemos antes fazer algo como:

Neste caso procuramos a linha da mesma forma que no caso anterior, no entanto, paramos caso cheguemos a uma linha com ÜTRANSACCAO¿já que esta é a a última linha da transacção então se a variável i for zero sabemos que a linha com Parque de estacionamentoïão existe nesta transacção.

Codificação e Testes

4.1 Alternativas, Decisões e Problemas de Implementação

Com o intuito de completar o programa extra E foi necessário estudar um pouco o formato de um ficheiro **HTML** com o qual não estávamos familiarizados a cem por cento. Inicialmente notamos que a informação era demasiada e ficava demasiado confusa ao colocar tudo numa só página, daí optou-se por dividí-la em diferentes secções que seriam apresentadas em diferentes páginas acedidas através de links. Isto significa que um input com mais dados pode gerar mais ficheiros **HTML**, no entanto, já que a intenção é apresentar a informação ao utilizador de forma simples, fácil e limpa, achou-se que os benefícios justificam o facto de o número de ficheiros criados variar com o input.

Mesmo dividindo em várias páginas, certos conjuntos de informações geraram páginas muito extensas, por exemplo uma página com todas as transacções de um extracto pode tornar-se bastante extensa. A nossa solução foi separar os dados referentes a cada transacção com texto com cor vermelha e um tamanho de letra maior que o restante texto chamando assim a atenção de que se iniciará uma nova transacção apartir daquela linha. Este texto atribui também um número sequencial a cada transacção a começar no um para fácilmente se encontrar determinada transacção relevante. Outra opção poderia ser organizar estes dados numa tabela ou então dividir as transacções por várias páginas e ter um link que permitiria avançar para a página seguinte.

4.2 Testes realizados e Resultados

Mostra-se agora imagens que demonstram o comando usado para cada programa e o output correspondente. **Programa A:**

4.1 Funcionamento do programa A

Programa B:

4.2 Funcionamento do programa B

Programa C:

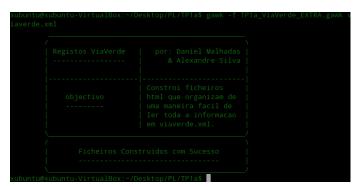
4.3 Funcionamento do programa C

Programa D:



4.4 Funcionamento do programa D

Programa E:



4.5 Funcionamento do programa E

Agora que se gerou os ficheiros do programa E podemos observá-los. O ficheiro base que se deve abrir primeiro é o ficheiro **index.html**. Aqui podemos escolher a funcionalidade que pretendemos, nesta versão apenas se pode consultar extractos.



4.5 ficheiro index.html

Ao carregar no link podemos observar todos os extractos. Como o ficheiro de input apenas contém um extracto então apenas vemos um na seguinte imagem, no entanto, se o input tivesse vários, eles apareceriam aqui.



Lista de Extractos Extracto 1: ID do Extracto: "011114056/08/2015" Mes de Emissao: Ago-2015 ID do Cliente: "514714936" NIF do Cliente: 987653210

Nome do Cliente:

PEDRO MANUEL RANGEL SANTOS HENRIQUES

Morada:

RUA XXX

Localidade:

BRAGA

Codigo Postal:

4715-012 BRAGA

Identificador:

"28876820811"

Matricula:

00-LJ-11

Referencia de Pagamento:

1234567

Consultar transaccoes deste extracto

Consultar viagens realizadas neste extracto

Consultar parques utilizados neste extracto

4.6 ficheiro de extractos

Temos agora 3 links para cada extracto. O primeiro mostra todas as transacções realizadas nesse extracto, o segundo todas as viagens (através das transacções que resultam de portagens) e o último os parques de estacionamento que visitou.

Olhamos agora para o link das transacções:



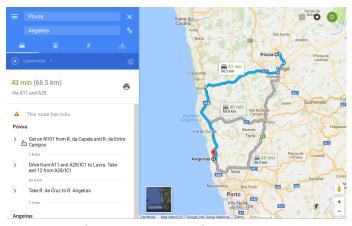
4.7 ficheiro de transacções

Aqui temos a lista de todas as transacções, sendo que as restantes se encontram por baixo, separadas pelo texto vermelho indicando o número de cada uma como referido na subsecção anterior. Olhamos agora para o link das viagens:



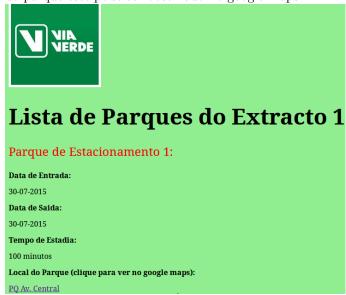
4.8 ficheiro de viagens

Aqui temos as viagens separadas também por texto vermelho indicando o número de cada uma. É ainda indicado o tempo de duração da mesma, sendo que no final da página temos o tempo total de todas as viagens. Clicando no percurso da viagem podemos observar o mesmo no google maps. Como exemplo temos o que aparece ao clicar no percurso da primeira viagem:



4.9 percurso da viagem 1 no google maps

Olhando agora por último para o link dos parques vemos os parques também divididos pelo texto vermelho. Temos também o tempo de estadia em cada parque e o tempo total passado em parques no fim da página. Ao clicar no local do parque este pode ser observado no google maps.



4.10 ficheiro de parques

Conclusão e Trabalho Futuro

Terminado o projeto entende-se que o mesmo cumpre com sucesso tudo o que era pretendido e adiciona ainda funcionalidades extra que são realmente relevantes para o contexto do problema. Em particular o programa E foi pensado com especial cuidado para que fosse realizado algo que teria algum propósito real e não só académico já que, poderia ser do interesse de alguém organizar a informação guardada num ficheiro **XML** de uma forma mais apelativa e fácil de entender para um humano. O facto de se ter escolhido uma interface com recurso a ficheiros **HTML** ao invés do terminal permitiria que estes fossem usados para, por exemplo, serem adicionados a um site onde o cliente poderia consultar os seus extractos.

Embora se tenha conseguido completar tudo o que se pretendeu, poder-se-ia ainda melhorar já que nunca se pode dar nenhum programa por realmente terminado ou perfeito. Poder-se-ia, por exemplo, organizar as transacções/viagens/parques por vários ficheiros, sendo que cada um apresentaria por exemplo dez itens de cada vez, dando depois a opção ao utilizador de navegar para um outro ficheiro com os próximos dez. Embora isto pudesse tornar as coisas mais organizadas o número de ficheiros ia ser muito maior por isso também teria as suas desvantagens.

Por fim, conclui-se que o projeto final ficou do nosso agrado tendo permitido que, com a prática, desenvolvessemos capacidades em GAWK que não tinhamos anteriormente. Desenvolvemos também uma melhor percepção do que realmente são expressões regulares e achamos estar agora mais aptos para perceber melhor o decorrer da matéria na unidade curricular em questão.

Apêndice A

Código do Programa

Programa A

```
BEGIN{
     print "\t _____"
     print "\t/

      print "\t/
      \\

      print "\t| ENTRADAS
      | por: Daniel Malhadas |"

      print "\t| ------
      | & Alexandre Silva |"

      print "\t|
      | "

      print "\t|-----|"

        print "\t|
        | Determinar quantas |"

        print "\t|
        objectivo | entradas em cada |"

        print "\t|
        ------ | dia, em cada mes e |"

        print "\t|
        | em cada ano. |"

     print "\t\\_____/"
     print "\t/
     #Contador de Entradas. Inicializado a O
     entradas = 0;
}
#Expressao regular que identifica uma entrada
/<DATA_ENTRADA>/{
     #Contador de data
           split ($0,data,">|<");</pre>
     datas[data[3]]++;
     #Contadores dia, mes e ano
      split (data[3],dma,"-");
           dia[dma[1]]++;
           mes[dma[2]]++;
           ano[dma[3]]++;
           entradas++;
}
END{
     print "\t|----"
```

```
print "\t| Entradas em cada data"
   print "\t|----"
   for (i in datas)
       print "\t| Na Data " i ": " datas[i] " entradas";
   print "\t|----"
   print "\t| Entradas em cada dia"
   print "\t|----"
   for (i in dia)
       print "\t| No dia " i ": " dia[i] " entradas";
   print "\t|----"
   print "\t| Entradas em cada mes"
   print "\t|----"
   for (i in mes) {
       if(length(i)==0) print "\t| No mes null: " mes[i] " entradas";
                 print "\t| No mes " i ": " mes[i] " entradas";
   }
   print "\t|----"
   print "\t| Entradas em cada ano"
   print "\t|----"
   for (i in ano) {
       if(length(i)==0) print "\t| No ano null: " ano[i] " entradas";
            print "\t| No ano " i ": " ano[i] " entradas";
   }
   print "\t|----"
   print "\t| Total Entradas: " entradas
   print "\t\\_____/"
}
Programa B
BEGIN{
   print "\t _____
   print "\t/

      print "\t| SAIDAS
      | por: Daniel Malhadas | "

      print "\t| -----
      | & Alexandre Silva | "

      print "\t|
      | "

   print "\t|-----|----|"
   print "\t| | Determinar que saidas |"
print "\t| objectivo | se realizaram, quantas |"
print "\t| ------ | de cada e quantas no |"
                             | total.
                                                     | "
   print "\t|
   print "\t\\_____/"
                                                      \\"
   print "\t/
   #Contador de saidas. Inicializado a 0
   saidas = 0;
}
#Expressao regular que identifica uma saida
/<SAIDA>/{
   #Contador de data
```

Programa C

```
BEGIN{
   print "\t ______
   print "\t/

      print "\t| Total Gasto
      | por: Daniel Malhadas |"

      print "\t| ------
      | & Alexandre Silva |"

      print "\t|
      | "

      print "\t|-----|"

   \\"
   print "\t/
    #Contador de total gasto. Inicializado a 0
   total_gasto = 0;
    #Contador de total desconto. Inicializado a 0
    total_desconto = 0;
    #Total pagamentos. Inicializado a 0
   pagamentos = 0;
}
#Expressao regular que identifica uma importancia
/<IMPORTANCIA>/{
    split ($0,importancia,">|<");</pre>
    sub(",",".",importancia[3]);
   total_gasto += importancia[3];
    gasto_agora = importancia[3];
   getline;
    split ($0,desconto,">|<");</pre>
    sub(",",".",desconto[3]);
    total_desconto += desconto[3];
    gasto_agora -= desconto[3];
    for(getline; (i=index($0,"DATA_DEBITO"))==0; getline);
    split($0,data,">|<");</pre>
    gasto_data[data[3]]+=gasto_agora;
    split(data[3],dma,"-");
    dia[dma[1]]+=gasto_agora;
   mes[dma[2]]+=gasto_agora;
    ano[dma[3]]+=gasto_agora;
   pagamentos++;
}
```

```
END{
   print "\t|----"
   print "\t| Pago em cada data"
   print "\t|----"
   for (i in gasto_data)
      print "\t| Na Data " i " pago " gasto_data[i];
   print "\t|----"
   print "\t| Pago em cada dia"
   print "\t|-----"
   for (i in dia)
      print "\t| No dia " i " pago " dia[i];
   print "\t|----"
   print "\t| Pago em cada mes"
   print "\t|----"
   for (i in mes) {
      print "\t| No mes " i " pago " mes[i];
   print "\t|----"
   print "\t| Pago em cada ano"
   print "\t|----"
   for (i in ano) {
      print "\t| No ano " i " pago " ano[i];
   }
   print "\t|----"
   print "\t| Total Pago : " total_gasto
   print "\t| Total Desconto : " total_desconto
   print "\t| Total Pagamentos: " pagamentos
   print "\t\\_____/"
}
Programa D
BEGIN{
   print "\t ______
   print "\t/
                                                    //"

      print "\t| Total Gasto
      | por: Daniel Malhadas |"

      print "\t| ------
      | & Alexandre Silva |"

      print "\t|
      | "

   print "\t|-----|-"
   print "\t|
   print "\t| objectivo | por dia/mes/ano apenas |"
print "\t| ------ | em parques. |"
                             | Determinar total gasto |"
   print "\t|
   print "\t\\______
                                                    \\"
   print "\t/
   #Contador de total gasto. Inicializado a 0
   total_gasto = 0;
   #Contador de total desconto. Inicializado a O
   total_desconto = 0;
```

```
#Total pagamentos. Inicializado a 0
   pagamentos = 0;
}
#Expressao regular que identifica uma importancia
/<IMPORTANCIA>/{
   split ($0,importancia,">|<");</pre>
   getline;
   split ($0,desconto,">|<");</pre>
   for(getline; ((i=index($0,"Parques de estacionamento"))==0) &&
               (index($0,"/TRANSACCAO>")==0);
       getline);
   if (i!=0){
       sub(",",".",importancia[3]);
       sub(",",".",desconto[3]);
      total_gasto += importancia[3];
       gasto_agora = importancia[3];
       total_desconto += desconto[3];
                 -= desconto[3];
       gasto_agora
      getline;
       split($0,data,">|<");</pre>
       gasto_data[data[3]]+=gasto_agora;
       split(data[3],dma,"-");
      dia[dma[1]]+=gasto_agora;
      mes[dma[2]]+=gasto_agora;
       ano[dma[3]]+=gasto_agora;
      pagamentos++;
   }
}
END{
   print "\t|-----"
   print "\t| Pago em cada data (apenas em Parques)"
   print "\t|-----"
   for (i in gasto_data)
      print "\t| Na Data " i " pago " gasto_data[i];
   print "\t|-----"
   print "\t| Pago em cada dia (apenas em Parques)"
   print "\t|-----"
   for (i in dia)
      print "\t| No dia " i " pago " dia[i];
   print "\t|-----"
```

```
print "\t| Pago em cada mes (apenas em Parques)"
   print "\t|-----"
   for (i in mes) {
       print "\t| No mes " i " pago " mes[i];
   print "\t|-----"
   print "\t| Pago em cada ano (apenas em Parques)"
   print "\t|-----"
   for (i in ano) {
      print "\t| No ano " i " pago " ano[i];
   print "\t|-----"
   print "\t| Total Pago apenas em Parques : " total_gasto
   print "\t| Total Desconto apenas em Parques : " total_desconto
   print "\t| Total Pagamentos apenas em Parques: " pagamentos
   print "\t\\_____/"
}
Programa E
BEGIN{
   print "\t ______
   print "\t/
   print "\t| Registos ViaVerde | por: Daniel Malhadas |"
print "\t| ------ | & Alexandre Silva |"
   print "\t|
   print "\t|-----|----|"
   print "\t| | Constroi ficheiros | "
print "\t| objectivo | html que organizam de | "
print "\t| ------ | uma maneira facil de | "
print "\t| | ler toda a informacao | "
   print "\t|
                             | em viaverde.xml. |"
   print "\t\\_____/"
   print "\t/
   #Constroi a base/home page dos registos ViaVerde
   base = "index.html" #ficheiro base
   extractos = "extractos.html" #ficheiro com os extractos
   transaccoes = "transaccoes.html" #ficheiro com transaccoes
   viagens = "viagens.html" #ficheiro com viagens
   parques = "parques.html" #ficheiro com parques
   print "<!DOCTYPE html>" > base;
   print "<html>" > base;
   print "<body>" > base;
   print "<body style=\"background-color:lightgreen\">" > base;
   print "<img src=\"viaverde.jpg\" width=\"200\" height=\"200\">" > base;
   print "<h1 style=\"color:darkgreen;font-size:300%;text-align:center\">Registos ViaVerde</h1>" > base;
   print "Funcionalidades:" > base;
   print "<a href=\"" extractos "\">Consultar extractos</a>" > base;
   print "</body>" > base;
   print "</html>" > base;
```

```
#Ficheiro base terminado
}
#Expressao regular que identifica uma importancia
/<EXTRACTO/{
#Identificar extracto
split ($0,idaux,">|<");</pre>
split (idaux[2],extracto_id,"=");
id = extracto_id[2];
getline;
split ($0,data,">|<");</pre>
data_emissao[id] = data[3];
getline;
split ($0,id_clienteaux,">|<");</pre>
split (id_clienteaux[2],id_cliente,"=");
clientes_id[id] = id_cliente[2];
getline;
split ($0,nif,">|<");</pre>
nifs[id] = nif[3];
getline;
split ($0,nome_cliente,">|<");</pre>
nome_clientes[id] = nome_cliente[3];
getline;
split ($0,morada,">|<");</pre>
moradas[id] = morada[3];
getline;
split ($0,localidade,">|<");</pre>
localidades[id] = localidade[3];
getline;
split ($0,cod_postal,">|<");</pre>
codigos_postais[id] = cod_postal[3];
getline;
getline;
split ($0,identificador_aux,">|<");</pre>
split (identificador_aux[2],identificador,"=");
identificadores[id] = identificador[2];
getline;
split ($0,mat,">|<");</pre>
matriculas[id] = mat[3];
getline;
split ($0,ref,">|<");</pre>
referencia_pagamento[id] = ref[3];
```

```
#Identificar transaccoes
t=0;
while(1){
    for(getline; (i=index($0,"<TRANSACCAO>"))==0 &&
                   (j=index($0,"</EXTRACTO>"))==0 ;getline);
    if (i==0) break;
    t++;
    getline;
    split ($0,t_data_e,">|<");</pre>
    t_datas_e[id,t] = t_data_e[3];
    getline;
    split ($0,horas_e,">|<");</pre>
    t_horas_entrada[id,t] = horas_e[3];
    getline;
    split ($0,e,">|<");
    t_entradas[id,t] = e[3];
    getline;
    split ($0,data_s,">|<");</pre>
    t_datas_s[id,t] = data_s[3];
    getline;
    split ($0,horas_s,">|<");</pre>
    t_horas_saida[id,t] = horas_s[3];
    getline;
    split ($0,s,">|<");
    t_{saidas[id,t]} = s[3];
    getline;
    split ($0,imp,">|<");</pre>
    sub(",",".",imp[3]);
    t_importancias[id,t] = imp[3];
    getline;
    split ($0,desc,">|<");</pre>
    sub(",",".",desc[3]);
    t_descontos[id,t] = desc[3];
    getline;
    split ($0,iva,">|<");</pre>
    t_{ivas[id,t]} = iva[3];
    getline;
    split ($0,0,">|<");</pre>
    t_operadores[id,t] = o[3];
    getline;
```

```
split ($0,tipo,">|<");
   t_tipos[id,t] = tipo[3];
   getline;
   split ($0,deb,">|<");
   t_datas_debito[id,t] = deb[3];
   getline;
   split ($0,cart,">|<");</pre>
   t_cartao[id,t] = cart[3];
}
n_transaccoes[id] = t;
}
END{
   n = 0;
   #Construir ficheiro de extractos
   print "<!DOCTYPE html>" > extractos;
   print "<html>" > extractos;
   print "<body>" > extractos;
   print "<body style=\"background-color:lightgreen\">" > extractos;
   print "<img src=\"viaverde.jpg\" width=\"200\" height=\"200\">" > extractos;
   print "<h1 style=\"font-size:300%\">Lista de Extractos</h1>" > extractos;
   for (i in data_emissao){
       n++;
       print "" > (n transaccoes);
       print "Extracto " n ":" > extractos;
       print "<b>ID do Extracto:</b>" > extractos;
       print "<p>"i "</p>" > extractos;
       print "<b>Mes de Emissao:</b>" > extractos;
       print ""data_emissao[i] "" > extractos;
       print "<b>ID do Cliente:</b>" > extractos;
       print ""clientes_id[i] "" > extractos;
       print "<b>NIF do Cliente:</b>" > extractos;
       print ""nifs[i] "" > extractos;
       print "<b>Nome do Cliente:</b>" > extractos;
       print ""nome_clientes[i] "" > extractos;
       print "<b>Morada:</b>" > extractos;
       print ""moradas[i] "" > extractos;
       print "<b>Localidade:</b>" > extractos;
       print ""localidades[i] "" > extractos;
       \label{lem:print "<b>Codigo Postal:</b>" > extractos;}
       print ""codigos_postais[i] "" > extractos;
       print "<b>Identificador:</b>" > extractos;
       print ""identificadores[i] "" > extractos;
       print "<b>Matricula:</b>" > extractos;
       print "" matriculas[i] "" > extractos;
       print "<b>Referencia de Pagamento:</b>" > extractos;
       print ""referencia_pagamento[i] "" > extractos;
```

```
print "<a href=\"" (n transaccoes) "\">Consultar transaccoes deste extracto</a>" > extractos;
print "" > extractos;
print "<a href=\"" (n viagens) "\">Consultar viagens realizadas neste extracto</a>" > extractos;
print "" > extractos;
print "<a href=\"" (n parques) "\">Consultar parques utilizados neste extracto</a>" > extractos;
#Construir ficheiro de transaccoes para este extracto
print "<!DOCTYPE html>" > (n transaccoes);
print "<html>" > (n transaccoes);
print "<body>" > (n transaccoes);
print "<body style=\"background-color:lightgreen\">" > (n transaccoes);
print "<img src=\"viaverde.jpg\" width=\"200\" height=\"200\">" > (n transaccoes);
print "<h1 style=\"font-size:300%\">Lista de Transaccoes do Extracto "n"</h1>" > (n transaccoes);
for (j=1;j<=n_transaccoes[id];j++){</pre>
   print "" > (n transaccoes);
   print "Transaccao " j ":" > (n transaccoes);
   print "<b>Data de Entrada:</b>" > (n transaccoes);
   print "" t_datas_e[i,j] "" > (n transaccoes);
   print "<b>Hora de Entrada:</b>" > (n transaccoes);
   print "" t_horas_entrada[i,j] "" > (n transaccoes);
   print "<b>Local de Entrada:</b>" > (n transaccoes);
   print "" t_entradas[i,j] "" > (n transaccoes);
   print "<b>Data de Saida:</b>" > (n transaccoes);
   print "" t_datas_s[i,j] "" > (n transaccoes);
   print "<b>Hora de Saida:</b>" > (n transaccoes);
   print "" t_horas_saida[i,j] "" > (n transaccoes);
   print "<b>Local de Saida:</b>" > (n transaccoes);
   print "" t_saidas[i,j] "" > (n transaccoes);
   print "<b>Importancia:</b>" > (n transaccoes);
   print "" t_importancias[i,j] "" > (n transaccoes);
   print "<b>Desconto:</b>" > (n transaccoes);
   print "" t_descontos[i,j] "" > (n transaccoes);
   print <b>IVA:</b>" > (n transaccoes);
   print "" t_ivas[i,j] "" > (n transaccoes);
   print "<b>Operador:</b>" > (n transaccoes);
   print "\p" t_operadores[i,j] "\p" > (n transaccoes);
   print "<b>Tipo de Transaccao:</b>" > (n transaccoes);
   print "" t_tipos[i,j] "" > (n transaccoes);
   print "<b>Data de Debito:</b>" > (n transaccoes);
   print "" t_datas_debito[i,j] "" > (n transaccoes);
   print "<b>Cartao Usado:</b>" > (n transaccoes);
   print "" t_cartao[i,j] "" > (n transaccoes);
print "</body>" > (n transaccoes);
print "</html>" > (n transaccoes);
#Ficheiro de transaccoes deste extracto terminado
#Construir ficheiro de viagens para este extracto
print "<!DOCTYPE html>" > (n viagens);
print "<html>" > (n viagens);
print "<body>" > (n viagens);
print "<body style=\"background-color:lightgreen\">" > (n viagens);
print "<img src=\"viaverde.jpg\" width=\"200\" height=\"200\">" > (n viagens);
```

```
print "<h1 style=\"font-size:300%\">Lista de Viagens do Extracto "n"</h1>" > (n viagens);
tempo_total = 0;
jj=0;
for (j=1;j<=n_transaccoes[id];j++){</pre>
   if (t_tipos[i,j]!="Portagens") continue;
   print "Viagem " jj ":" > (n viagens);
   print "<b>Data de Entrada:</b>" > (n viagens);
   print "" t_datas_e[i,j] "" > (n viagens);
   print "<b>Data de Saida:</b>" > (n viagens);
   print "" t_datas_s[i,j] "" > (n viagens);
   if (t_datas_e[i,j]!="null" && t_datas_s[i,j]!="null"){
       print "<b>Tempo de Viagem:</b>" > (n viagens);
       split(t_datas_e[i,j],d,"-");
       split(t_horas_entrada[i,j],h,":");
       de = mktime(d[3] " " d[2] " " d[1] " " h[1] " " h[2] " 00");
       split(t_datas_s[i,j],d,"-");
       split(t_horas_saida[i,j],h,":");
       ds = mktime(d[3] " " d[2] " " d[1] " " h[1] " " h[2] " 00");
       tempo_total += ((ds - de)/60);
       print "((ds - de)/60)" minutos" > (n viagens);
   print "<b>Percurso Realizado (clique para ver no google maps):</b>" > (n viagens);
   print "<a href=\"https://www.google.pt/maps/dir/"t_entradas[i,j]"/"t_saidas[i,j]"/\"/>De " t_entr
print "rolor:red;font-size:160%;\">Tempo Total de Viagem "tempo_total" minutos
print "</body>" > (n viagens);
print "</html>" > (n viagens);
#Ficheiro de viagens deste extracto terminado
#Construir ficheiro de parques para este extracto
print "<!DOCTYPE html>" > (n parques);
print "<html>" > (n parques);
print "<body>" > (n parques);
print "<body style=\"background-color:lightgreen\">" > (n parques);
print "<img src=\"viaverde.jpg\" width=\"200\" height=\"200\">" > (n parques);
print "<h1 style=\"font-size:300\%\">Lista de Parques do Extracto "n"</h1>" > (n parques);
tempo_total = 0;
jj=0;
for (j=1;j<=n_transaccoes[id];j++){</pre>
   if (t_tipos[i,j]!="Parques de estacionamento") continue;
   jj++;
   print "Parque de Estacionamento " jj ":" > (n parques)
   print "<b>Data de Entrada:</b>" > (n parques);
   print "" t_datas_e[i,j] "" > (n parques);
   print "<b>Data de Saida:</b>" > (n parques);
   print "" t_datas_s[i,j] "" > (n parques);
   if (t_datas_e[i,j]!="null" && t_datas_s[i,j]!="null"){
       print "<b>Tempo de Estadia:</b>" > (n parques);
       split(t_datas_e[i,j],d,"-");
       split(t_horas_entrada[i,j],h,":");
       de = mktime(d[3] " " d[2] " " d[1] " " h[1] " " h[2] " 00");
       split(t_datas_s[i,j],d,"-");
       split(t_horas_saida[i,j],h,":");
```

```
ds = mktime(d[3] " " d[2] " " d[1] " " h[1] " " h[2] " 00");
           tempo_total+=((ds - de)/60);
           print ""((ds - de)/60)" minutos<math>" > (n parques);
        }
        print "<b>Local do Parque (clique para ver no google maps):</b>" > (n parques);
        print "color:red;font-size:160%;\">Tempo Total em Parques "tempo_total" minutos
     print "</body>" > (n parques);
     print "</html>" > (n parques);
     #Ficheiro de parques deste extracto terminado
  }
  print "</body>" > extractos;
  print "</html>" > extractos;
  #Ficheiro de extractos terminado
  print "\t\\_____/"
}
```