

Processamento de Linguagens (3º ano de Curso)

Trabalho Prático 2 (GAWK)

Relatório de Desenvolvimento

José Pinto
(A81317)

Luís Correia
(A81141)

Pedro Barbosa
(A82068)

28 de Abril de 2019

Resumo

Este relatório é referente ao segundo trabalho prático da UC de Processamento de Linguagens que tem como objetivo aumentar a capacidade de escrever *Expressões regulares* (ER) para descrição de padrões de frases com o intuito de filtrar ou transformar textos. Para isso vai se recorrer a ferramenta *GAWK* para a realização de programas para a filtragem de textos. Neste relatório são descritas as várias etapas do desenvolvimento dos pontos pedidos.

Conteúdo

1	Introdução	2
2	Análise e Especificação	3
2.1	Descrição do problema	3
3	Concepção/desenho da Resolução	4
3.1	Resposta às alíneas	4
3.1.1	Alínea a)	4
3.1.2	Alínea b)	5
3.1.3	Alínea c)	5
3.1.4	Alínea d)	6
3.1.5	Extras	7
4	Codificação e Testes	10
4.1	Alternativas, Decisões e Problemas de Implementação	10
4.2	Testes realizados e Resultados	11
4.2.1	Alínea a)	11
4.2.2	Alínea b)	12
4.2.3	Alínea c)	12
4.2.4	Alínea d)	13
4.2.5	Extras	16
5	Conclusão	19

Capítulo 1

Introdução

O enunciado deste segundo trabalho prático continha 5 exercícios, dos quais se devia proceder a resolução de um exercício.

A escolha do exercício baseou-se no número de aluno mais baixo do grupo. Com isso, o exercício resolvido para este primeiro trabalho foi o número 2 (Processador de Processos de Formação).

Capítulo 2

Análise e Especificação

2.1 Descrição do problema

O exercício prático 2 (Processador de Processos de Formação) consiste na análise e filtragem de um ficheiro que contém informação sobre a criação e gestão de cursos de formação profissional.

Na primeira alínea (alínea a) foi pedido que se procedesse a uma limpeza do ficheiro, isto é, que se retirassem linhas vazias extras e as linhas que continham os campos todos vazios. Para além disso, sempre que o campo *Estado* estivesse vazio, este devia passar a ter o valor de **NIL**.

Uma vez realizada a alínea a), o ficheiro a ser filtrado e processado para as alíneas seguintes era o resultado desta mesma alínea.

A segunda alínea consiste numa contagem dos registos que apresentam um código numérico, apresentando para estes o código, o título, a descrição e as notas.

Na alínea c) o objetivo é identificar os diferentes tipos de processos que existem e contar o número de processos relativos a cada tipo.

Por último, foi pedido um grafo desenhado em DOT que relacionasse o código com todos os diplomas jurídico-administrativos usados.

Para além disto, ainda foi sugerida a inclusão de outros processamentos para além destes.

Capítulo 3

Concepção/desenho da Resolução

3.1 Resposta às alíneas

3.1.1 Alínea a)

Como mencionado anteriormente, nesta alínea é preciso limpar o ficheiro, criando um segundo ficheiro processado. É necessário remover linhas vazias extra e as que possuem os campos todos vazios. Também é preciso indicar que o valor de todos os registos que possuem *Estado* vazio passa a ser *NIL*.

Depois de uma análise ao ficheiro com informação relativa a cursos de formação profissional, concluiu-se que o que separava os campos era o carácter ";" (ponto e vírgula), passando então este a ser o *Field Separator* (FS). Também se observou que o que separava os registos era uma linha vazia. Sendo assim, definiu-se o RS (*Record Separator*) como "\n\n+". Com os delimitadores definidos, procedeu-se a filtragem do ficheiro dado.

O programa *AWK* processa todos os registos que não correspondam à expressão regular "{34}", neste caso isto é, que não sejam um registo (uma linha) que contém todos os campos vazios.

Para todos esses registos, verifica-se se o primeiro campo está vazio. Em caso afirmativo, coloca-se o valor **NIL** nesse campo imprimindo de seguida o resto do registo. Em caso negativo, imprime-se para o ficheiro de saída todo o registo. No fim de imprimir um registo adiciona-se uma quebra de linha (\n) para que no ficheiro resultado o *record separator* continue a ser uma linha vazia.

Na figura 3.1 apresenta-se o programa *AWK* relativo a esta alínea.

```
BEGIN {
    FS="; "
    RS="\n\n+"

    $0 !~/{34}/ {
        if($1 == "") print "NIL" $0 "\n" >> "processado.txt"
        else print $0 "\n" > "processado.txt"
    }
```

Figura 3.1: Programa AWK da alínea a).

3.1.2 Alínea b)

Nesta alínea pretende-se contar o número de registos com um Código numérico e mostrar para estes o respetivo Código, título, descrição e notas.

Para esta alínea o ficheiro a filtrar é o resultado do processamento efetuado na alínea a). O *Field Separator* continua a ser o caracter ”;”. Porém, o *Record Separator* passou a ser ”;\n\n+” pois pode-se observar pelo ficheiro que existem linhas vazias no meio de registos.

Antes ainda de se proceder ao processamento do ficheiro, inicia-se um contador (*conta*) a zero. Este vai contar o número de registos que possuem um código.

O programa AWK começa a processar depois do registo 1 para que este não seja tido em conta, visto que apenas contém a informação de todos os campos de cada registo. Também só são contados os registos onde o campo *Código* não é vazio.

Para cada registo é impresso o correspondente código, título, descrição e as notas.

No final do processamento é imprimido o número de registos que contém código.

Na figura 3.2 tem-se o programa AWK para esta alínea.

```
BEGIN {
    FS=";";
    RS=";\n\n+";
    conta = 0;
}

#Começa depois do registo 1 porque este apenas contém a informação dos campos de cada registo.
NR > 1 && $2 !~/"/ {
    print "\nCÓDIGO: " $2 "\nTÍTULO: " $3 "\nDESCRIÇÃO: " $4 "\nNOTAS: " $5;
    conta ++;
}

END {
    print "\n\nExistem " conta " registos que contém um Código";
}
```

Figura 3.2: Programa AWK da alínea b).

3.1.3 Alínea c)

Na alínea c) o objetivo é identificar os diferentes tipos de processos e para cada tipo calcular o número de processos relativos a este.

O *Field Separator* e o *Record Separator* continuam a ser igual à alínea anterior.

Sendo assim, processa-se todos os registos depois do primeiro e para cada tipo encontrado incrementa-se uma unidade na posição desse tipo no array *contaTipos*. Esta acção é possível ser observada na figura 3.3.

No final do processamento é imprimido para cada tipo, o número de processos que são desse tipo. De referir que os processos que não possuem tipo também são contados e que são apresentados como *Sem tipo*.

```

BEGIN {
    FS=";";
    RS=";\n\n+"
}

NR > 1 {
    contaTipos[$10]++;
}

END {
    print "Tipo de processo -> número de processos correspondentes";
    for(tipo in contaTipos){
        if(tipo != ""){
            print tipo " -> " contaTipos[tipo];
        }
        else {
            print "Sem tipo -> " contaTipos[tipo];
        }
    }
}

```

Figura 3.3: Programa AWK da alínea c).

3.1.4 Alínea d)

A alínea d) consiste em desenhar um grafo em DOT relacionando os códigos de cada ação de formação com todos os diplomas jurídico-administrativos usados.

O *Field Separator* e o *Record Separator* continuam a ser igual à alínea anterior.

O resultado deste programa *AWK* é um ficheiro .dot, posteriormente dando origem ao grafo pedido.

Antes do início da filtragem, é imprimido no ficheiro .dot a designação que indica que se vai iniciar um grafo e algumas especificações para que o grafo seja de fácil visualização.

Em seguida, processa-se cada registo, imprimindo para o ficheiro .dot todos os códigos e os diplomas jurídico-administrativos relativo a um código de um processo.

Existem dois tipos de diplomas jurídico-administrativos, os REF e os complementares. Observando o ficheiro resultado da alínea a) é possível observar que os diplomas jurídico-administrativos REF num registo são separados pelo carácter "#". Sendo assim, usa-se a função *split* para colocar num array todos os diplomas para que se possam associar estes ao seu registo. De notar que foi necessário o uso da função *substr* para o primeiro diploma pois este vem com uma aspa(") no início, ignorando assim esta. Também é preciso referir que apenas se percorre o array com todos os diplomas a partir da posição 2 até à penúltima pois o primeiro diploma é um caso à parte que foi explicado anteriormente e o último não é um diploma, mas sim a aspa final que delimita os diplomas.

Por último, associa-se também os diplomas jurídico-administrativos complementares aos seus códigos.

No fim do processamento, imprime-se para o ficheiro .dot a chaveta para indicar o fim do grafo. Com isto, é possível gerar o grafo em DOT.

Na figura 3.4 é possível observar o programa *AWK* relativo a esta alínea d). É possível observar que existem algumas linhas que servem para definir a cor de cada nó do grafo, para que se possa distinguir pelas cores qual o tipo do nó. Neste caso, se a cor do nó for verde indica que é um código, se for vermelho indica que é um diploma jurídico-administrativos REF e se for azul claro é um diplomas jurídico-administrativos complementar.

```
BEGIN {
    FS=";";
    RS=";\n\n+";

    #rankdir = LR para que o grafo se faça da esquerda para a direita
    print "digraph{\n    rankdir=LR;" > "d.dot"
}

NR > 1 {
    print " \\\n" $2 "\\n" " "[style=filled, color=\\".3 .4 .8\\n\";" > "d.dot";

    #Separa todos os diplomas, estes estão entre #
    n = split($8, diplomasREF, "#\n");
    #Todos os diplomas estão entre aspas, logo isto faz com que se separe o primeiro diploma da primeira aspa.
    primeiroREF = substr(diplomasREF[1], 2);
    if(primeiroREF != "") {
        print " \\\n" primeiroREF "\\n" " "[style=filled, color=\\".1.0 .6 1.0\\n\";" > "d.dot";
        print " \\\n" $2 "\\n" -> " \\\n" primeiroREF "\\n\";" > "d.dot";
    }
    #No ciclo só se percorre do campo 2 ao penúltimo, pois o ultimo é uma aspa também
    for(i = 2; i < n; i++){
        if(diplomasREF[i] != "\n"){
            print " \\\n" diplomasREF[i] "\\n" " "[style=filled, color=\\".1.0 .6 1.0\\n\";" > "d.dot";
            print " \\\n" $2 "\\n" -> " \\\n" diplomasREF[i] "\\n\";" > "d.dot";
        }
    }

    #Só para os registos que tem diplomas complementares
    if($9 != ""){
        print " \\\n" substr($9, 1, length($9)-1) "\\n" " "[style=filled, color=\\".5 .5 1.0\\n\";" > "d.dot";
        print " \\\n" $2 "\\n" -> " \\\n" substr($9, 1, length($9)-1) "\\n\";" > "d.dot";
    }
}

END {
    print "}" > "d.dot"
}
```

Figura 3.4: Programa AWK da alínea d).

3.1.5 Extras

Para além dos pontos pedidos, foram realizados alguns extras como complemento ao exercício. Sendo assim, vão-se listar os programas *AWK* extra que foram feitos.

De referir que para todos eles o FS (*field separator*) é ";" e o RS (*record separator*) é ";\n\n+"

- **Metadados**

Para uma melhor compreensão e facilidade em perceber qual a posição de cada campo, fez-se um programa que processa a primeira linha do ficheiro fornecido, linha esta que possui informação sobre o ficheiro. Para cada campo, imprime para um ficheiro denominado de "meta.txt" a posição do campo e o que representa esse campo.

Este programa é apresentado na figura 3.5.

```
NR == 1 {
    for(i = 1; i <= NF; i++){
        print i "->" $i > "meta.txt"
    }
}
```

Figura 3.5: Programa AWK para metadados.

- **Dimensão qualitativa de processos**

Os processos possuem um campo que indica qual a dimensão qualitativa destes. Com base nisso, contou-se os vários valores para esse campo, indicando para cada o número de processos com essa dimensão qualitativa. Isto é possível observar na figura 3.6.

```
NR > 1 {
    if($18 != "") dimensao[$18]++;
}
```

Figura 3.6: Programa AWK para dimensão qualitativa.

No final do processamento, é apresentado o resultado. Esta ação (figura 3.7 é incluída na secção *END*.

```
for(dim in dimensao){
    print dimensao[dim] " processos com dimensão qualitativa " dim;
}
```

Figura 3.7: Programa AWK para dimensão qualitativa.

- **Grafo identificando donos de processos**

Cada processo possui um campo que identifica o dono deste. Para este extra, o objetivo é criar um grafo que apresente para cada dono os seus processos.

Antes de iniciar o processamento do ficheiro, é necessário imprimir para um ficheiro denominado "donos.dot" a cláusula que indica o início de um grafo, *"digraph{ \n rankdir=LR;" > "donos.dot"*.

Depois desta ação feita, percorre-se o ficheiro de formações a partir do segundo registo, imprimindo para o ficheiro "donos.dot" as associações entre um dono e um código.

Não esquecer que no fim do processamento é necessário incluir uma chaveta no final do ficheiro indicando que acabou o grafo.

Este programa é apresentado na figura 3.8.

```

NR > 1 {
    n = split($12, donos, "#");
    if(n <= 2){
        if( donos[1] != "") print "\" donos[1] \" -> \" \" $2 \" \" > \"donos.dot\";
    }
    else{
        dono1 = substr(donos[1], 2);
        if(dono1 != "")
            print "\" dono1 \" -> \" \" $2 \" \" > \"donos.dot\";
        for(i = 2; i < n; i++){
            if(donos[i] != ""){
                print "\" donos[i] \" -> \" \" $2 \" \" > \"donos.dot\";
            }
        }
    }
}

```

Figura 3.8: Programa AWK para grafo de donos.

- **Prazo de conservação administrativa de cada processo**

Por último, fez-se um programa que apresenta para todos os códigos o respetivo prazo de conservação administrativa.

Este programa percorre todos os registos e para estes guarda num array o prazo do processo. Tal esta apresentado na figura 3.9.

```

NR > 1 {
    if($20 != ""){
        prazos[$2] = $20;
    }
}

```

Figura 3.9: Programa AWK para prazos de conservação administrativa.

Acabando o processamento do ficheiro, apresenta-se para todos os códigos qual o prazo respetivo Isto é feito na secção *END*.

```

print "\nCódigo de processo -> Prazo de conservação admnistrativa: "
for(i in prazos){
    print i " -> " prazos[i];
}

```

Figura 3.10: Programa AWK para prazos de conservação administrativa.

Capítulo 4

Codificação e Testes

4.1 Alternativas, Decisões e Problemas de Implementação

Na concepção deste trabalho prático encontrou-se um entrave na filtragem do ficheiro fornecido. Este ficheiro tinha a extensão *.csv* e quando se aplicava um programa *AWK* a este ficheiro, não se obtia o resultado pretendido.

Para resolver este problema converteu-se a extensão do ficheiro para *.txt*. Uma vez executada esta ação, o mesmo programa *AWK* aplicado ao ficheiro, agora *.txt*, já executava e dava como *output* o resultado pretendido.

Outro problema mencionado acima na explicação dos exercícios foi o caso de existirem linhas em branco no meio de registos, tal como é apresentado na figura 4.1.

```
AP#
Grupo IP#;"DGES#
AP#;"Comunicar#
Iniciar#;"150.20.101#
150.20.300#
150.40.001#
250.20.001#
350.30.001#
450.30.502#
500.10.301#
750.10.001#
750.10.300#
750.20.301#
750.20.600#
750.30.602#
800.20.001#
800.20.002#
800.20.500#

";"Planeamento estratégico de atividades#
Avaliação de atividades#
Realização de operações estatísticas oficiais#
Registo biográfico de trabalhadores#
Cobrança de receitas e pagamento de despesas#
Emissão de declarações comprovativas#
Realização de fiscalização#
Seleção e seriação para ingresso no ensino ou formação#
Processamento dos dados cadastrais de alunos ou formandos#
Distribuição de atividades de ensino ou formação#
Realização de atividades de ensino ou formação#
Reconhecimento, creditação e validação de competências e qualificações#
Realização de estudos de investigação básica#
Execução de projetos de investigação aplicada#
Desenvolvimento experimental e aplicacional#;"Sucessão (sucessor)#
Síntese (sintetizado)#
Síntese (sintetizado)#
Síntese (sintetizado)#
```

Figura 4.1: Registo com linha em branco.

Para que tal não fosse considerado como dois registos, mas sim um registo tal como é, alterou-se o RS(*record separator*) para ”;\n\n+”. Deste modo, vai poder existir linhas em branco a meio de registos.

4.2 Testes realizados e Resultados

Nesta secção mostram-se alguns testes feitos (valores introduzidos) e os respectivos resultados obtidos:

4.2.1 Alínea a)

O programa *AWK* desta alínea cria um ficheiro, denominado de *processado.txt*, que é o resultado de retirar as linhas vazias extra e as que possuem os campos todos vazios. Para além disso, sempre que o *Estado* está vazio é colocado **NIL** no seu lugar.

Na figura 4.2 é possível observar um excerto do ficheiro *processado.txt*.

```
1 Estado;Código;Título;Descrição;Notas de aplicação;Exemplo NA;Notas de exclusão;Diplomas jurídico-administrativos REF;Diplomas jurídico-administrativos complementar;Tipo de processo;Processo transversal (S
2 /N);Dono do processo;Participante no processo;Tipo de intervenção do participante;Código do processo relacionado;Título do processo relacionado;Tipo de relação entre processos;Dimensão qualitativa do
3 processo;Uniformização do processo;Prazo de conservação administrativa;Nota ao PCA;Justificação PCA;Forma de contagem do PCA normalizada;Forma de contagem do prazo de cons. adm.;Destino
4 final;Justificação DF;Notas;:::::::
5 NIL;750;PRESTAÇÃO DE SERVIÇOS DE ENSINO E FORMAÇÃO;"Relativo à prestação de serviços no domínio da educação/ensino/qualificação da população, independentemente da idade ou do contexto (escolar,
6 académico, profissional ou outro) - inclui a educação pré-escolar, o ensino básico e secundário, o ensino superior, a educação extraescolar e todos os cursos de formação, de
7 qualificação profissional e valorização permanente, em qualquer área do conhecimento,
8 ";Inclui as ações de formação para a qualificação dos recursos humanos da administração pública, bem como a realização de estágios.#
9 Aplicável tanto às entidades que prestam os serviços referidos como às que solicitam a prestação dos mesmos.#";"As ações de sensibilização, congressos, seminários e outros encontros científicos ou
10 profissionais, devem ser considerados em ""Dinamização e comunicação institucional/ Execução de atividades de dinamização da sociedade"" (900.10).#
11 "::::::::::::::::::::::::::
12 NIL;750.10;Gestão do aluno/formando;Compreende as atividades relacionadas com apresentação de candidaturas, seleção de candidatos, matrículas, inscrições, transferências de alunos/formandos entre cursos,
13 turmas ou estabelecimentos de ensino, e ainda as atividades relacionadas com a orientação vocacional do aluno/formando, e/ou o seu encaminhamento para percurso ou ensino especial.;Quando aplicável, estas
14 atividades incidem igualmente sobre estagiários.#::::::::::::::::::::::::::::::::::
15 NIL;750.10.001;Seleção e seriação para ingresso no ensino ou formação;"Avaliação da capacidade para a frequência de curso ou de ação de formação, bem como a realização das tarefas inerentes ao concurso
16 ou admissão, designadamente as provas específicas, e à satisfação de pré-requisitos (realização de provas de aptidão física, funcional ou vocacional).
17 Inicia com a candidatura a curso ou inscrição/encaminhamento para formação e termina com a publicação da lista de graduação final ou lista de admitidos a formação.
18 Inclui realização de provas de ingresso, aplicação de fórmulas e classificações mínimas, seriação dos candidatos, elaboração de pautas finais e respetiva homologação.";"Candidatos selecionados ou
19 encaminhados para a formação profissional#
20 Candidatos ao ingresso no âmbito da escolaridade obrigatória e do ensino regular profissional#
21 Concursos especiais (maiores de 23 anos)#
22 Seriação dos candidatos ao curso Curso de Estudos Avançados em Gestão Pública (CEAGP)#
23 Bolseiro dos Países Africanos de Língua Oficial Portuguesa (PALOP) e Comunidade dos Países de Língua Portuguesa (CPLP)#
24 Curso institucional#
25 Reingresso, mudança de curso e transferência#
26 Pós-graduações#
27 Curso de Defesa Nacional#
28 Formação ministrada pelo Centro de Formação de Funcionários de Justiça, para acesso e progressão na carreira de oficial de justiça#
29 Formação para acesso e progressão na carreira de guarda prisional#
30 Seriação para ingresso de recluso ao ensino básico e secundário e ensino superior#
31 Seriação para ingresso de jovem internado em centro educativo ao ensino básico e secundário#
32 Concurso de Ingresso na Formação Inicial - Tribunal Administrativo e Fiscal e Tribunais Judiciais#
33 Candidatos identificados para reforço de conhecimentos na área da Propriedade Industrial (PI) segundo pré-requisitos#
34 Estágio para Ingresso na Carreira de Investigação Criminal da Polícia Judiciária#
35 Formação ministrada no âmbito da medicina legal e ciências forenses#
36 Formação ministrada pelo Centro de Formação do IRN#
37 Formação para acesso às carreiras de conservador, notário e escriturário#
38 Estágio para ingresso na carreira de conservador e notário#";"A elaboração de calendários das provas de ingresso deve ser considerada em 150.20.101 - Planeamento estratégico de atividades#
39 A reclamação de resultados da seleção deve ser considerada em 500.40.001 - Processamento de reclamações em atos administrativos, ou em 500.40.500 - Processamento de recursos hierárquicos, ou, ainda, em
40 500.40.501 - Processamento de recursos administrativos especiais#
41 O recurso que evolua para os Tribunais deve ser considerado em 600.30.550 - Ação administrativa declarativa comum#
42 As atividades relacionadas com a seleção e seriação de recluso para acesso a ação de formação devem ser consideradas em 600.40.303 - Acompanhamento da execução de medida tutelar educativa não privativa
43 da liberdade#
44 As atividades relacionadas com a seleção e seriação de jovem em centro educativo para acesso a ação de formação devem ser consideradas em 600.40.100 - Execução de medida de internamento em centro
45 educativo#";"Lei 45/2013#
46 Lei 45/2012#
47 DL 92/2011#
48 Lei 60/2011#
```

Figura 4.2: Ficheiro *processado.txt*.

Na figura 4.3 é possível observar o final do ficheiro *processado.txt*. É possível observar que não possui registos onde os campos estão todos vazios, caso que acontecia com o ficheiro *formacao* fornecido.

```
969 NIL:750.30.602.01;Reconhecimento, creditação e validação de competências e qualificações: verificação das condições;"Inicia com a verificação e análise do percurso formativo e termina com relatório
preliminar.
970 Inclui elaboração do portefólio individual que explicita e organiza as evidências das competências adquiridas.";;;;;;;;;;;;;5;"#Critério legal:
971 Prazo prescricional do procedimento contraordenacional ([DL 433/82], Ilícito de mera ordenação social, artº 27;
972 [DL 4/2015], CPA, artº 168 - Condicionaisismos aplicáveis à anulação administrativa - "" 1 - Os atos administrativos podem ser objeto de anulação administrativa no prazo de seis meses, (...) desde que não
tenham decorrido cinco anos, a contar da respetiva emissão
973 #Critério de utilidade administrativa:
974 Prazo decorrente da necessidade de consulta deste PN para apuramento de responsabilidade em sede de ação disciplinar (PN 500.20.002 - Realização de inquérito, 500.20.003 - Realização de sindicância) e de
fiscalização (500.10.301 - Realização de fiscalização)e para processamento de novos plano).";Data de conclusão do procedimento;Data de encerramento do processo;CP;"#Critério de densidade informacional:
975 Enumeração e tipificação dos casos recuperável através das estatísticas.
976 Esta síntese não permite recuperar o particular e individual, mas apenas o geral.
977 A aplicação de um critério de amostragem aleatória sobre este PN permitirá conhecer em detalhe alguns casos e obter um subconjunto representativo de todas as características da população-alvo.
978 Toma ainda como base o facto da imprevisibilidade dos estudos nesta área e a impossibilidade atual de conservar toda a produção.
979 A constituição de uma amostra significativa do universo de onde é retirada, mediante a aplicação de métodos estatísticos com uma margem de erro controlada, será suficiente para representar o universo.
980 A grelha para seleção da amostragem encontra-se no final desta página.
981 Extra amostra poderão ser conservadas os fat files ou os processos referentes a casos com maior impacto social.";;;;;;;;;
982
983 NIL:750.30.602.02;Reconhecimento, creditação e validação de competências e qualificações: atribuição de equivalência ou reconhecimento;"Inicia com análise do relatório preliminar da análise do percurso
formativo e termina com a definição da qualificação.
984 Inclui reuniões e elaboração de pareceres.";;;;;;;;;;;;;5;"#Critério legal:
985 Prazo prescricional do procedimento contraordenacional ([DL 433/82], Ilícito de mera ordenação social, artº 27;
986 [DL 4/2015], CPA, artº 168 - Condicionaisismos aplicáveis à anulação administrativa - "" 1 - Os atos administrativos podem ser objeto de anulação administrativa no prazo de seis meses, (...) desde que não
tenham decorrido cinco anos, a contar da respetiva emissão
987 #Critério de utilidade administrativa:
988 Prazo decorrente da necessidade de consulta deste PN para apuramento de responsabilidade em sede de ação disciplinar (PN 500.20.002 - Realização de inquérito, 500.20.003 - Realização de sindicância) e de
fiscalização (500.10.301 - Realização de fiscalização) e para processamento de novos plano).";Data de conclusão do procedimento;Data de encerramento do processo;C;"#Critério de densidade informacional:
989 Informação não recuperável noutro PN.";;;;;;;;;
990
991
```

Figura 4.3: Final do ficheiro processado.txt.

4.2.2 Alínea b)

Na figura 4.4 é possível observar o resultado do programa *AWK* relativo à alínea b).

O resultado apresenta o código, título, descrição e notas de todos os registos que apresentem *Código*.

```
gawk -f b.awk processado.txt

CÓDIGO: 750
TÍTULO: PRESTAÇÃO DE SERVIÇOS DE ENSINO E FORMAÇÃO
DESCRIÇÃO: "Relativo à prestação de serviços no domínio da educação/ensino/qualificação da população, independentemente da idade ou do contexto (escolar, académico, profissional ou outro) - inclui a educação pré-escolar, o ensino básico e
secundário, o ensino superior, a educação extraescolar e todos os cursos de formação, de
qualificação profissional e valorização permanente, em qualquer área do conhecimento.
"
NOTAS:

CÓDIGO: 750.10
TÍTULO: Gestão do aluno/formando
DESCRIÇÃO: Compreende as atividades relacionadas com apresentação de candidaturas, seleção de candidatos, matrículas, inscrições, transferências de alunos/formandos entre cursos, turmas ou estabelecimentos de ensino, e ainda as atividades
relacionadas com a orientação vocacional do aluno/formando, e/ou o seu encaminhamento para percurso ou ensino especial.
NOTAS:

CÓDIGO: 750.10.001
TÍTULO: Seleção e seriação para ingresso no ensino ou formação
DESCRIÇÃO: "Avaliação da capacidade para a frequência de curso ou de ação de formação, bem como a realização das tarefas inerentes ao concurso ou admissão, designadamente as provas específicas, e à satisfação de pré-requisitos (realização
de provas de aptidão física, funcional ou vocacional).
Inicia com a candidatura a curso ou inscrição/encaminhamento para formação e termina com a publicação da lista de graduação final ou lista de admitidos a formação.
Inclui realização de provas de ingresso, aplicação de fórmulas e classificações mínimas, seriação dos candidatos, elaboração de pautas finais e respetiva homologação."
NOTAS:

CÓDIGO: 750.10.001.01
TÍTULO: Seleção e seriação para ingresso no ensino ou formação: seleção
DESCRIÇÃO: "Inicia com a candidatura / encaminhamento.
Inclui apresentação de elementos instrutórios, verificação dos pré-requisitos, aplicação dos métodos para a seleção, quando necessário, como a realização de provas de ingresso (gerais ou específicas) e respetiva avaliação, elaboração e pu
blicitação de lista provisória de admitidos e excluídos, reclamações.
Termina com publicitação de lista de admitidos e excluídos."
NOTAS:
E sintetizada no
PN 150.40.500 - Recolha e tratamento de dados estatísticos de apoio à gestão."

CÓDIGO: 750.10.001.02
TÍTULO: Seleção e seriação para ingresso no ensino ou formação: seriação
DESCRIÇÃO: "Inicia-se com a aplicação dos métodos para a ordenação dos candidatos.
Inclui a elaboração de listas ou pautas definitivas.
Termina com a publicitação.
"
NOTAS:

CÓDIGO: 750.10.002
TÍTULO: Processamento de matrículas ou inscrições no ensino ou em formação
DESCRIÇÃO: "Realização ou renovação de matrícula em cursos ou inscrição em ações de formação.
Inicia com o pedido de acesso ou ingresso e termina com a entrega de comprovativo de matrícula ou inscrição.
Inclui verificação de dados de identificação e validação da existência dos requisitos necessários para efeito de matrícula ou inscrição.
"
NOTAS:
PN 750.20.600 - Realização de atividades de ensino ou formação."
```

Figura 4.4: Resultado do programa AWK da alínea b).

4.2.3 Alínea c)

Para a alínea c) era pretendido mostrar todos os tipos diferentes e calcular o número de processos por cada tipo.

Na figura 4.5 é possível observar o resultado do programa *AWK* da alínea c).

```
gawk -f c.awk processado.txt
Tipo de processo -> número de processos correspondentes
Sem tipo -> 12
PC -> 12
PE -> 5
```

Figura 4.5: Resultado do programa AWK da alínea c).

Pelo resultado, conclui-se que existem 12 processos que não possuem um tipo, outros 12 que são do tipo **PC** e por fim 5 processos com tipo **PE**.

4.2.4 Alínea d)

O objetivo desta alínea é relacionar num grafo cada código de um processo com os seus diplomas jurídico-administrativos. Para isso foi feito um programa *AWK* que gera um ficheiro em DOT.

Parte do ficheiro em DOT gerado é o apresentado na figura 4.6. É possível observar que se colocou o grafo a desenhar da esquerda para a direita através da linha *randir=LR*. Outro aspecto a mencionar é a atribuição de cores aos nodos, para que se possam diferenciar melhor uns dos outros.

```

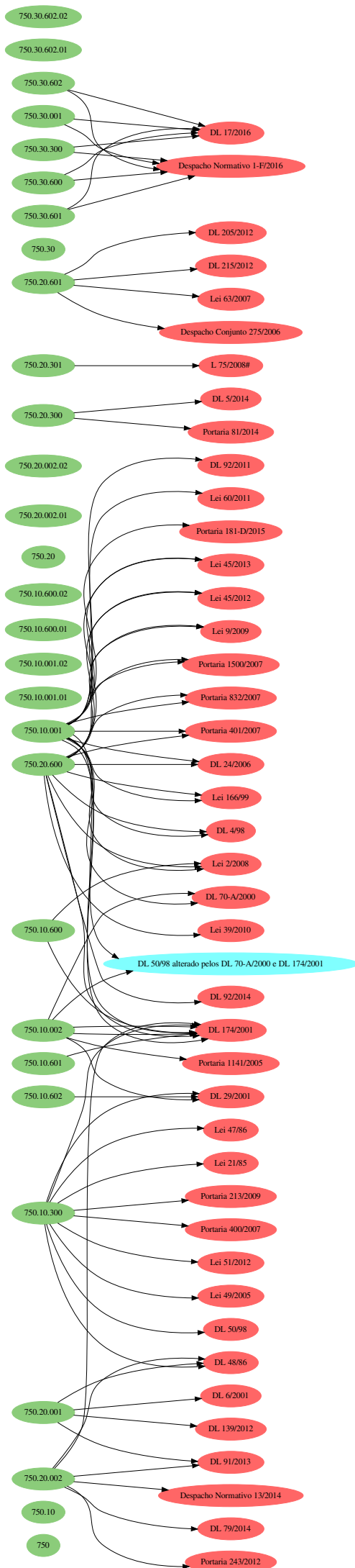
digraph{
  rankdir=LR;
  "750" [style=filled, color=".3 .4 .8"];
  "750.10" [style=filled, color=".3 .4 .8"];
  "750.10.001" [style=filled, color=".3 .4 .8"];
  "Lei 45/2013" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "Lei 45/2013";
  "Lei 45/2012" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "Lei 45/2012";
  "DL 92/2011" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "DL 92/2011";
  "Lei 60/2011" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "Lei 60/2011";
  "Portaria 181-D/2015" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "Portaria 181-D/2015";
  "Lei 9/2009" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "Lei 9/2009";
  "Lei 2/2008" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "Lei 2/2008";
  "Portaria 1500/2007" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "Portaria 1500/2007";
  "Portaria 832/2007" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "Portaria 832/2007";
  "Portaria 401/2007" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "Portaria 401/2007";
  "DL 24/2006" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "DL 24/2006";
  "DL 174/2001" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "DL 174/2001";
  "Lei 166/99" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "Lei 166/99";
  "DL 70-A/2000" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "DL 70-A/2000";
  "DL 174/2001" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "DL 174/2001";
  "DL 4/98" [style=filled, color="1.0 .6 1.0"];
  "750.10.001" -> "DL 4/98";
  "DL 50/98 alterado pelos DL 70-A/2000 e DL 174/2001#" [style=filled, color=".5 .5 1.0"];
  "750.10.001" -> "DL 50/98 alterado pelos DL 70-A/2000 e DL 174/2001#";
  "750.10.001.01" [style=filled, color=".3 .4 .8"];
  "750.10.001.02" [style=filled, color=".3 .4 .8"];
  "750.10.002" [style=filled, color=".3 .4 .8"];
  "Portaria 1141/2005" [style=filled, color="1.0 .6 1.0"];
  "750.10.002" -> "Portaria 1141/2005";
  "DL 174/2001" [style=filled, color="1.0 .6 1.0"];
  "750.10.002" -> "DL 174/2001";
  "DL 29/2001" [style=filled, color="1.0 .6 1.0"];
  "750.10.002" -> "DL 29/2001";
  "DL 70-A/2000" [style=filled, color="1.0 .6 1.0"];
  "750.10.002" -> "DL 70-A/2000";
  "DL 174/2001" [style=filled, color="1.0 .6 1.0"];
}

```

Figura 4.6: Excerto do ficheiro DOT resultante do programa AWK.

Uma vez gerado este ficheiro em DOT r penas produzir o grafo. Na esta afigura abaixo é possível observar o grafo (em DOT) produzido a partir do resultado do programa AWK.

De referir que os nodos que possuem cor verde representam o *Código* do processo, os com cor vermelha os *Diplomas jurídico-administrativos REF* e os com cor azul clara os *Diplomas jurídico-administrativos complementar*.



4.2.5 Extras

- **Metadados**

O objetivo deste programa era a criação de um ficheiro de metadados. O ficheiro é apresentado na figura 4.7.

```
1->Estado
2->Código
3->Título
4->Descrição
5->Notas de aplicação
6->Exemplo NA
7->Notas de exclusão
8->Diplomas jurídico-administrativos REF
9->Diplomas jurídico-administrativos complementar
10->Tipo de processo
11->Processo transversal (S/N)
12->Dono do processo
13->Participante no processo
14->Tipo de intervenção do participante
15->Código do processo relacionado
16->Título do processo relacionado
17->Tipo de relação entre processos
18->Dimensão qualitativa do processo
19->Uniformização do processo
20->Prazo de conservação administrativa
21->Nota ao PCA
22->Justificação PCA
23->Forma de contagem do PCA normalizada
24->Forma de contagem do prazo de cons. adm.
25->Destino final
26->Justificação DF
27->Notas
28->
29->
30->
31->
32->
33->
34->
```

Figura 4.7: Ficheiro de metadados.

- **Dimensão qualitativa de processos**

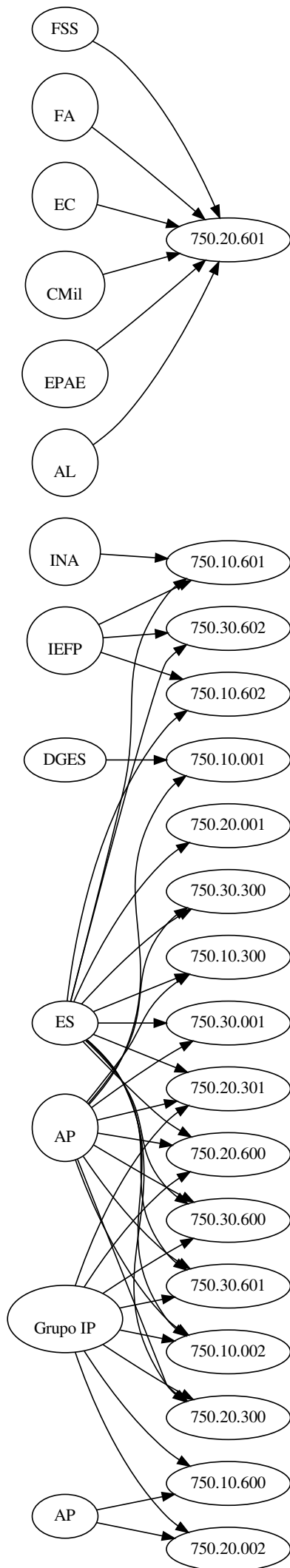
Neste programa o resultado é a apresentação do número de processos para cada dimensão qualitativa. O resultado é apresentado na figura 4.8.

```
11 processos com dimensão qualitativa Elevada
6 processos com dimensão qualitativa Média
```

Figura 4.8: Resultado de dimensão qualitativa de processos.

- **Grafo identificando donos de processos**

Para este extra o objetivo era a concepção de um grafo em DOT que mostre quais os códigos de cada dono. O resultado é apresentado no seguinte grafo.



- **Prazo de conservação administrativa de cada processo**

O resultado do programa que apresenta os prazos de conservação administrativa de cada processo é apresentado na figura 4.9. O resultado apresenta primeiro o Código do processo e depois o prazo correspondente.

```
Código de processo -> Prazo de conservação admnistrativa:
750.30.001 -> 5
750.30.300 -> 5
750.20.001 -> 5
750.30.600 -> 5
750.20.300 -> 6
750.10.600.01 -> 10
750.30.601 -> 5
750.20.301 -> 5
750.10.600.02 -> 1
750.10.001.01 -> 5
750.30.602.01 -> 5
750.20.600 -> 10
750.20.002.01 -> 5
750.10.300 -> 60
750.10.002 -> 5
750.10.001.02 -> 5
750.30.602.02 -> 5
750.20.601 -> 10
750.20.002.02 -> 5
750.10.601 -> 1
750.10.602 -> 5
```

Figura 4.9: Resultado de Prazos de cada processo.

Capítulo 5

Conclusão

Este trabalho permitiu um maior contacto com a ferramenta *GAWK*, contribuindo para um aprofundamento e solidificação do uso de expressões regulares (ER), usadas para o desenvolvimento de *Processadores de Linguagens Regulares*.

Uma vez realizado este trabalho pode-se concluir que o objetivo final foi alcançado, pois realizou-se todos os pontos propostos, ainda havendo espaço para alguns extras.