

Universidade do Minho

LEI — Licenciatura de Engenharia Informática

Processamento de Linguagens

# Trabalho prático 1

Paulo Araujo - a58925, Orlando Costa - a67705, Rui Oliveira - a67661



Braga, 25 de março de 2015

## **Resumo**

Este relatório descreve a resolução de um conjunto de exercícios propostos, que consistem no desenvolvimento de programas na linguagem C com o auxílio de geradores de filtros de texto, como o Flex. Para cada problema é realizada uma breve análise sobre o trabalho efetuado, as decisões que lideraram o seu desenvolvimento e as estruturas implementadas, assim como uma explicação do seu funcionamento.

Os problemas resolvidos consistem no desenvolvimento de filtros de texto que:

- processa um ficheiro XML com descrições de fotografias e gerar um álbum HTML.
- processa de um ficheiro XML anotado com tags Enamex e gera páginas HTML apresentando as "pessoas", "países", "cidades"e organizações nele identificadas.
- processa vários ficheiros de texto, compostos por letras de canções, e gera documentos em LATEX para cada uma delas.

Para cada problema é apresentado o código em linguagem C e as expressões regulares desenvolvidas, sendo estes suportados por exemplos e devidos resultados.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>3</b>
<b>2</b>	<b>Museu da Pessoa — tratamento de fotografias</b>	<b>4</b>
2.1	Analise e Especificação . . . . .	4
2.2	Implementação . . . . .	4
2.2.1	Estrutura de dados . . . . .	4
2.2.2	Filtro de Texto . . . . .	5
2.2.3	Funcionamento . . . . .	6
2.3	Testes realizados . . . . .	6
<b>3</b>	<b>Processamento de Entidades Nomeadas</b>	<b>8</b>
3.1	Analise e Especificação . . . . .	9
3.2	Implementação . . . . .	9
3.2.1	Estrutura de dados . . . . .	9
3.2.2	Filtro de Texto . . . . .	10
3.2.3	Funcionamento . . . . .	11
3.3	Testes realizados . . . . .	12
<b>4</b>	<b>Processamento de ficheiros com Canções</b>	<b>14</b>
4.1	Analise e Especificação . . . . .	14
4.2	Implementação . . . . .	14
4.2.1	Estrutura de dados . . . . .	14
4.2.2	Filtro de Texto . . . . .	15
4.2.3	Funcionamento . . . . .	15
4.3	Testes realizados . . . . .	16
<b>5</b>	<b>Conclusão</b>	<b>17</b>
<b>6</b>	<b>Anexos</b>	<b>18</b>
6.1	Museu da Pessoa — tratamento de fotografias . . . . .	18
6.1.1	Filtro de Texto . . . . .	18
6.1.2	Estrutura de dados . . . . .	19
6.1.3	Cabeçalho ficheiro C . . . . .	24
6.1.4	Testes . . . . .	24
6.2	Processamento de Entidades Nomeadas (Enamex) . . . . .	41
6.2.1	Filtro de Texto . . . . .	41
6.2.2	Estrutura de dados . . . . .	44

6.2.3	Cabeçalho ficheiro C . . . . .	47
6.2.4	Testes . . . . .	48
6.3	Processamento de ficheiros com Canções . . . . .	54
6.3.1	Filtro de Texto . . . . .	54
6.3.2	Estrutura de dados . . . . .	55
6.3.3	Cabeçalho ficheiro C . . . . .	59
6.3.4	Testes . . . . .	59

# Capítulo 1

## Introdução

O presente projeto enquandra-se na unidade curricular de Processamento de Linguagens do curso de Licenciatura em Engenharia Informática da Universidade do Minho. O projeto pretende aumentar as capacidades com as expressões regulares, desenvolvendo processadores de linguagens regulares utilizando o gerador de filtros de texto Flex. Para isso foram selecionados 3 exercícios dentro de um grupo de 8 exercícios, são eles: *2.1 Museu da Pessoa — tratamento de fotografias*, *2.2 Processamento de Entidades Nomeadas (Enamex)* e *2.5 Processamento de ficheiros com Canções*.

## Capítulo 2

# Museu da Pessoa — tratamento de fotografias

Neste problema é pretendido que, a partir de um ficheiro *XML*, seja criado um filtro de texto capaz de interpretar informação relativa às entrevistas feitas para construção o Museu da Pessoa. Com toda essa informação, deverá ser criado um álbum em *HTML* possuidor de um índice, ordenado alfabeticamente, contendo todos os nomes de pessoas presentes no álbum. Além disso, cada elemento do índice deverá estar referenciado para a página que contém todas as fotos da respetiva pessoa. Relativamente às fotos, essas deverão estar ordenadas cronologicamente e a descrição da foto deverá ser o seu título/cabeçalho.

### 2.1 Analise e Especificação

Após a análise do que é pedido no enunciado e de alguns dos *Datasets*, foi possível verificar qual a informação essencial a retirar do ficheiro *XML*. Além disso, como não é regra que as fotos e pessoas presentes no ficheiro *XML* estejam já na ordem pretendida, será necessário que toda a informação seja armazenada em estruturas de dados. O nome do ficheiro *HTML* resultante, poderá ser dado como argumento e caso não seja, por omissão será *AlbumGerado.html*. Além deste ficheiro, serão gerados tantos ficheiros *HTML* quantas as pessoas presentes no álbum. Esses ficheiros serão denominados numericamente (1.html,2.html,...) à medida que novos nomes são encontrados. Os ficheiros anteriormente referidos apenas serão gerados no final da leitura e filtragem de ficheiro *XML*.

### 2.2 Implementação

#### 2.2.1 Estrutura de dados

De forma a dar seguimento ao que foi especificado na secção 2.1, foi necessária a criação de uma estrutura de dados que sirva de suporte aos dados recolhidos. Sendo assim, optamos por criar uma estrutura denominada `photo_node` que terá o formato de uma árvore binária e servirá para

armazenar toda a informação relativa a cada descrição de fotos encontrada no ficheiro *XML*. Nesta estrutura serão guardados o nome do ficheiro que contém a foto, a data e o local em que esta foi tirada, a sua descrição e o nome das pessoas nela presentes. A inserção de fotos nesta árvore será ordenada relativamente à sua data. Além desta estrutura, também criamos outra, denominada **person\_node**, com o objetivo que esta guarde toda a informação acerca das pessoas presentes no álbum. Esta estrutura tem o formato de uma lista ligada e guardará informação como o nome da pessoa, assim como o nome do ficheiro da sua página *HTML* e terá ainda um apontador para uma estrutura **photo\_node** em que será armazenada toda a informação relacionada com as suas fotos. Finalmente, criamos uma estrutura **Album** que apenas conterá um apontador para uma estrutura **person\_node**, onde estará a informação relativa às pessoas nele presente, e um contador que servirá para contar o número de pessoas presentes no álbum. O contador servirá de auxílio na criação dás páginas *HTML* numeradas referidas na secção [2.1](#).

### 2.2.2 Filtro de Texto

Um dos objetivos deste trabalho prático é a utilização de geradores de filtro de texto, como o *Flex*. Sendo assim, foi criado um ficheiro Flex que permite encontrar determinados padrões de expressões regulares e executar uma determinada ação para cada uma delas.

No ficheiro referido, podemos encontrar algumas instruções em linguagem C como a inclusão de ficheiros de cabeçalhos ( `headers.h`) e a declaração de variáveis. De seguida, são definidas as expressões regulares e as respetivas ações que se pretendem realizar no caso da identificação positiva do referido padrão:

```
\<foto[ \t]+[a-zA-Z]+=\".*\\" A partir da análise do ficheiro XML exemplificado no enunciado, foi possível verificar que a descrição de uma foto começa com o nome do ficheiro que a contém da seguinte forma: <foto ficheiro="ficheiro.jpg">. Sendo assim, quando este padrão é encontrado, é inicializado um photo_node com o nome encontrado entre as aspas. O nome é obtido retirando da expressão apenas o que se encontra à frente da primeira ocorrência de aspas e atrás da ocorrência seguinte.
```

```
\<quando[ \t]+[a-zA-Z]+=\\"[0-9]{4}(.|-)[0-9]{2}(.|-)[0-9]{2}\\> De seguida pretendemos encontrar a data em que a foto foi tirada. A partir do exemplo do enunciado, verificamos que a data é descrita da seguinte forma: <quando data="1961-01-15"/>. Sendo assim, pretendemos encontrar todas as expressões no formato referido, com a possibilidade que a data esteja separada por pontos em vez de traços. A obtenção da data faz-se retirando apenas o que se encontra à frente da primeira ocorrência de aspas.
```

```
\<quem>[ \t\na-zA-Z\u,\\\"0-9;:]+ Um padrão essencial a encontrar é o nome das pessoas presentes na foto. Esses nomes devem estar descritos da seguinte forma:
```

```
<quem>Ana de Lourdes de Oliveira Chamine; Antonio Oliveira Machado</quem>  
Nesta fase, optamos por não separar ainda os nomes por tokens porque será mais útil fazê-lo apenas quando a descrição da foto estiver completa. Sendo assim, apenas retiramos tudo o que esteja entre '>' e '<', e adicionamos à estrutura que contém a descrição da foto.
```

```
\<onde>[ \t\n0-9a-zA-Z\u.,;:\\"]+< Continuando com a análise do exemplo do enunciado, verificamos que uma foto pode ter descrito o local em que esta foi tirada:
```

```
<onde>Casa Machado, Afurada, Vila Nova de Gaia</onde>
```

Ora, quando este padrão é encontrado, só é necessário guardar o que está entre '>' e '<', assim como no caso anterior.

\<facto>[ \t\n0-9a-zA-ZÀ-Ù.,;:\"]+< Finalmente, o último campo necessário para a descrição da foto é o facto que esta representa: <facto> Os noivos cortam o bolo de casamento</facto>. O que é essencial retirar neste caso é o mesmo que nos casos anteriores, ou seja, apenas o que se encontra entre '›' e '‹'.

\<\foto> Sempre que é encontrado o padrão que finaliza a descrição de uma foto, </foto>, é necessário adicionar o `photo_node` criado a todos os respetivos `person_node` das pessoas que se encontram na descrição da foto. É nesta fase que os nomes das pessoas presentes na foto é separado e identificado. Finalmente adiciona-se a descrição da foto a todas as pessoas nela presente.

. | \n Este padrão apenas é utilizado para indicar que sempre que é encontrado qualquer outro byte ou \n, deve ser ignorado.

### 2.2.3 Funcionamento

Antes do filtro de texto ser aplicado ao ficheiro *XML* é invocada a função `init` que inicializa uma variável `static Album`. Inicializada esta variável, é possível aplicar o filtro de texto. À medida que a informação do ficheiro *XML* é filtrada, são invocadas funções que tratam e guardam na estrutura a informação recolhida. Sempre que é encontrada uma expressão que defina o início de uma descrição de uma foto, a informação relativa ao nome da foto é recolhida e é invocada a função `initPhoto` que trata de alocar o espaço necessário para a informação que virá a ser recolhida posteriormente. Do mesmo modo, sempre que é encontrada uma expressão que identifique a data, a localização, a descrição ou as pessoas presentes na foto, são invocadas funções capazes de tratar e armazenar essa informação. As funções referidas são `setDate`, `setLoc`, `setFact` e `setWho`, respetivamente. Sempre que não seja possível identificar uma destas características anteriores, estas ficam com os seus valores por defeito, ou seja, caso uma foto não tenha a tag <quem>, o seu campo correspondente na estrutura ficará preenchido com "Desconhecidos". Finalmente, quando é encontrada a expressão que define o final da descrição de uma foto, é invocada a função que trata de inserir a informação da nova foto na estrutura `Album`. Depois de tratada toda a informação, são invocadas funções auxiliares que criam e preenchem o(s) ficheiro(s) *HTML* necessários.

## 2.3 Testes realizados

Estão documentados neste secção 3 testes realizados ao autómato, utilizando como input os ficheiros que se encontram em anexo: [6.1.4](#), [6.1.4](#), [6.1.4](#) e [6.1.4](#).

### Teste nº 1

Neste teste podemos verificar que apenas existe uma página criada devida à presença de apenas uma pessoa nas fotos.

### **Teste nº 2**

Neste teste podemos verificar que as fotos não contêm descrição, visto que o ficheiro XML não contem o campo <facto>".

### **Teste nº 3**

Neste teste podemos verificar que um dos items do indice encontra-se incompleta, por não ser possível a leitura de um caracter.

### **Teste nº 4**

Teste exemplificado no enunciado. O nome das pessoas presentes na foto encontram-se separados por ":"; pode-se verificar que ambas as páginas respetivas contêm as fotos em comum.

Possíveis exemplos de aplicação deste filtro de texto:

- cat\ <inputFile>\ |\  
• cat\ <inputFile>\ |\  
play\

## Capítulo 3

# Processamento de Entidades Nomeadas

Entidades nomeadas são "elementos atómicos em texto" pertencentes a categorias predefinidas tais como nomes de pessoas, organizações, localizações, quantidades, etc. Assim, Processamento de Entidades Nomeadas(PEN) é a tarefa de identificar estas entidades. Embora as categorias das entidades nomeadas serem predefinidas, existem várias opiniões sobre que categorias deve ser consideradas entidades nomeadas e quão abrangentes estas categorias devem ser. Por convenção, tags "*ENAMEX*" são utilizadas para nomes, tags "*NUMEX*" são utilizadas para entidades numéricas, e tags "*TIMEX*" são utilizadas para entidades temporais.

Neste exercício iremos apenas processar entidades com a tag "*ENAMEX*", na forma:

- <ENAMEX TYPE="PERSON">Francisco de Vilela Barbosa</ENAMEX>  
(Pessoa)
- <ENAMEX TYPE="LOCATION" SUBTYPE="COUNTRY">Portugal</ENAMEX>  
(Localização, País)
- <ENAMEX TYPE="LOCATION" SUBTYPE="CITY">Rio de Janeiro</ENAMEX>  
(Localização, Cidade)
- <ENAMEX TYPE="ORGANIZATION">Universidade do Minho</ENAMEX>  
(Organização)

Como exercício extra iremos também abordar as entidades na forma:

- <ENAMEX TYPE="LOCATION"> Santo Novo </ENAMEX>  
(Localização não específica)

Todas as outras tags irão ser ignoradas.

O processamento de entidades nomeadas, apesar de ser aparentemente uma tarefa simples, enfrenta um dado numero de desafios. As entidades podem tornar-se difíceis de encontrar, e uma vez encontradas, difíceis de classificar. Localizações e nomes de pessoas podem ser as mesmas, e seguir estilos similares de formatação.

### 3.1 Analise e Especificação

Uma breve leitura do problema permite-nos entender algumas das funcionalidades necessárias, sendo estas definidas como:

- Necessidade de ordenação e não repetição na listagem de pessoas(alínea a):
  - A alínea A do problema requer a listagem de todas as pessoas identificadas, sem repetições. Este informação refere-se então às tags do tipo:  
`<ENAMEX TYPE="PERSON">...</ENAMEX>`.  
Por forma a armazenar e ordenar adequadamente toda a informação acerca das Pessoas, é necessária a utilização de uma estrutura capaz de suportar esta informação.
- Listar os países e cidades marcadas (alínea b):
  - Apesar de não estar especificado no enunciado, o grupo propôs uma implementação na qual seria possível associar cidades a certos países. Compreendemos que este tipo de implementação, para grande parte dos casos, não é viável e pode tornar a informação apresentada incoerente. No entanto, no intuito de aprender e aumentar o desafio proposto, decidimos que cidades mencionadas após países e antes de pontos finais pertenciam a esses países. Esta informação é também armazenada na estrutura implementada.
- Listar as organizações (alínea c):
  - Similarmente à alínea A, a alínea C requer a listagem de todas as organizações identificadas. Esta informação refere-se então às tags do tipo:  
`<ENAMEX TYPE="ORGANIZATION">...</ENAMEX>`  
E está implementada de forma similar à listagem de pessoas.
- Apresentar os resultados em formato HTML:
  - Por forma a visualizar facilmente os resultados do processamento do texto, estes são apresentados em formato HTML através da implementação de funções capazes de transformar a informação contida nas estruturas em documentos de texto com o formato requerido.

### 3.2 Implementação

#### 3.2.1 Estrutura de dados

Com o intuito de cumprir todos os requisitos estruturais definidos anteriormente, foi desenvolvida uma estrutura de dados única capaz de armazenar todos os dados necessários. Assim, escolhemos implementar uma árvore binária de procura, na qual os nodos possuem a informação a guardar sobre a forma de array de caracteres. A escolha desta estrutura facilita a ordenação alfabética dos diversos nomes que possamos processar, e é de implementação relativamente simples. Creemos ser superior a outras estruturas tais como listas ligadas cuja implementação, apesar de mais

simples, torna-se mais complexa quando é necessária a ordenação dos seus elementos ( $O(N)$ ) e as tabelas de hash cuja implementação é mais complexa, sendo que para elevadas quantidades de dados possuem ainda a necessidade de reHashing e garbage collection.

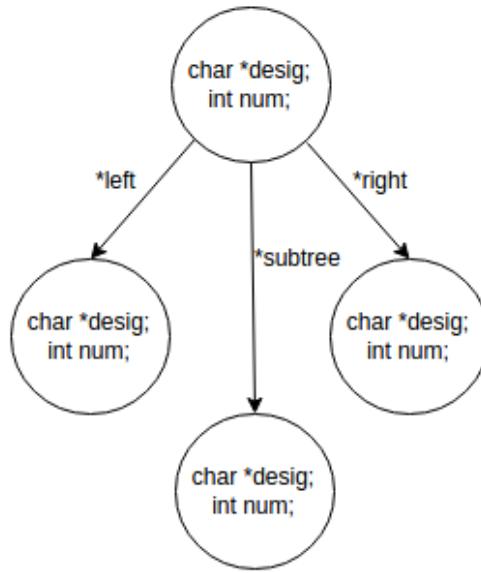


Figura 3.1: Representação gráfica da estrutura de dados

Para casos em que apenas é necessário o armazenamento da entidade, sem qualquer tipo de associação, uma árvore com dois apontadores (esquerda e direita) e com a capacidade de armazenar a informação (array de caracteres) bastaria para abordar todos os casos. No entanto, de forma a armazenar a possível associação entre os países e as suas cidades, modificamos a árvore previamente referida, e adicionamos-lhe um apontador extra, que poderá ser visto como o apontador para a raiz de uma nova árvore, constituída por todas as cidades que pertencem a um dado país. Como curiosidade, adicionamos também um inteiro em cada nodo que servirá como contador para todas as ocorrências de um dado elemento. Todo o código referente a esta estrutura encontra-se implementado no ficheiros "tree.c" e "tree.h" e pode ser consultado em anexo.

### 3.2.2 Filtro de Texto

De forma a processar as tags pertencentes a pessoas/organizações, foram criadas expressões regulares capazes de identificar essas tags. Assim, foram implementadas as seguintes expressões regulares para pessoas/organizações, assim como abreviaturas que facilitam a sua leitura e compreensão:

- pessoas: {enamex}{person}{pal}{eclose}
  - organizações: {enamex}{org}{pal}{ecclose}
  - localizações gerais: {enamex}{loc}{pal}{ecclose}
  - cidades: {enamex}{loc}{subcity}{pal}{ecclose}

sendo que as abreviaturas mencionadas correspondem a:

- Palavra  
`{pal} [a-zA-Z0-9ç-ñâ-û ]+`
- Tag Enamex  
`{enamex} \<[ \t]*(?i:enamex)[ \t]+(?i:type)=`
- Tag de fecho  
`{eclose} \<[ \t]*\/(?i:enamex)[ \t]*\>`
- Elemento "Organization"  
`{org} \"[ \t]*(?i:organization)[ \t]*\"[ \t]*\>`
- Elemento "Person"  
`{person} \"[ \t]*(?i:person)[ \t]*\"[ \t]*\>`

De notar a capacidade das expressões regulares identificarem tags definidas tanto em letra maiúscula como minúscula, sendo também tolerantes à quantidade de espaços ou tabs presentes entre elementos destas. De forma a implementar a capacidade de associar cidades a um dado país, foram utilizados "operadores de contexto", de forma a que caso seja detectado uma tag correspondente a um país, o analisador entre no contexto não exclusivo (%s)country, e associa as seguintes tags correspondentes a cidades ao país em causa. Caso seja detetado um ponto final, o analisador lexico abandona esse contexto e continua a processar em contexto geral.

**Nota:** Esta foi uma funcionalidade assumida, pelo que poderá nem sempre ter sucesso e fazer as corretas associações.

Assim as expressões regulares correspondentes a localizações/países/cidades foram definidas na forma:

- países:  
`{enamex}{loc}{subcountry}{pal}{eclose}`
- ponto final no contexto "country":  
`<country>`
- cidades no contexto "country":  
`<country>{enamex}{loc}{subcity}{pal}{eclose}`

sendo que as abreviaturas mencionadas correspondem a:

- Elemento "Country"  
`{subcountry} (?i:subtype)=\"[ \t]*(?i:country)[ \t]*\"[ \t]*\>`
- Elemento "City"  
`{subcity} (?i:subtype)=\"[ \t]*(?i:city)[ \t]*\"[ \t]*\>`

Estas expressões devem encontrar-se no topo de todas as outras, devido à precedência que possuem sobre elas.

### 3.2.3 Funcionamento

No cabeçalho do ficheiro flex são declarados todos os apontadores referentes às estruturas onde irá ser armazenada a informação. Existe um apontador para cada tipo de estrutura, nomeadamente: pessoas, países, cidades, organizações e outras localizações. São também declarados

dois apontadores para arrays de caracteres que irão auxiliar o processamento das expressões capturadas. Após início do programa, é efetuada a chamada ao analisador léxico, responsável por capturar os dados referentes às expressões definidas. Cada vez que este efetua uma captura, estes dados são processados, sendo inseridos na estrutura correspondente, sendo que caso necessário lhes são retirados os espaços em branco que possam ter antes e depois da seu conteúdo, por forma a evitar inconsistência de dados. Após o término da leitura do input em questão, todos os dados presentes nas estruturas são escritos nos ficheiros HTML correspondentes, estando estes interligados através de hiperligações (tags `<a href=>`). Cada estrutura (tipo de entidade) é escrita em ficheiro através da função `treeToHTML` (implementada no ficheiro `tree.c`), responsável por receber como parâmetros um apontador para estrutura e um identificador de ficheiro (previamente declarado), e transferir a informação para o ficheiro no formato adequado.

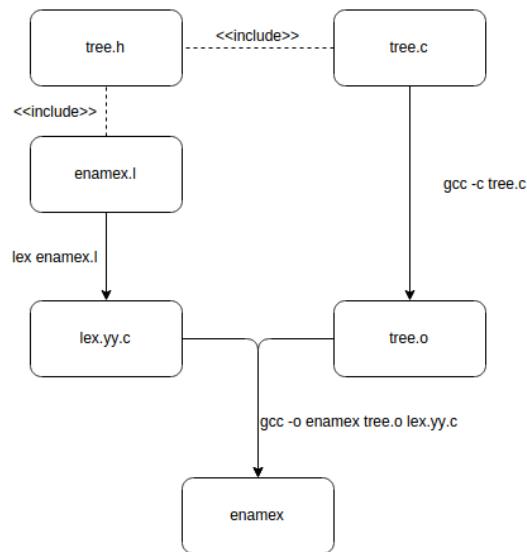


Figura 3.2: Componentes do programa

### 3.3 Testes realizados

Estão documentados neste secção 3 testes realizados ao autómato, utilizando como input os ficheiros que se encontram em anexo: [6.2.4](#), [6.2.4](#) e [6.2.4](#).

## Teste nº 1

Este focado na funcionalidade de associar diversas cidades a um dado país. Como é possível verificar em anexo, o resultado é uma conexão entre "Portugal" e algumas das suas cidades.

**Teste n° 2**

Teste exemplificado no enunciado. Processado de acordo com o esperado (exemplo da página referente aos locais).

**Teste n° 3**

Teste obtido de um caso real, e focado nas tags de organizações. Resultado correto.

## Capítulo 4

# Processamento de ficheiros com Canções

Neste problema era pretendido que fosse criado um filtro de texto capaz de interpretar ficheiros com letras de músicas, e fosse gerado um ficheiro *latex* para cada música encontrada, contendo a informação processada. Ainda existe a particularidade de que cada ficheiro com músicas poder conter mais do que uma música, sendo que neste caso devem ser criados 2 ficheiros *latex*.

### 4.1 Analise e Especificação

Existem várias questões que são deixadas em aberto no enunciado que irão ser especificadas nesta secção. O programa lê do *standard input* e os nomes dos ficheiros *latex* que irão ser gerados podem ser recebidos como argumento. Caso estes não sejam especificados, assumem um título numerado, começando em 0 até à n-ésima musica interpretada. Uma vez que não se sabe a ordem pela qual os cabeçalhos se apresentam nos ficheiros a serem interpretados, o mais seguro será guardar todas as musicas em memoria e apenas imprimir para o ficheiro *latex* após o fim de cada música. Após a análise dos *Datasets* verificou-se a existência de campos no cabeçalho que não são utilizados pelo programa, devendo estes ser ignorados. Durante a análise foi também verificada a existência de anotações em algumas músicas que serviriam para apresentar as pautas, pelo que optamos por ignorar estas marcas e tentar apenas imprimir a letra da música do ficheiro *latex*. Existe ainda outro cuidado na criação do ficheiro *latex* que é apresentação na música de caracteres especiais no *latex*.

### 4.2 Implementação

#### 4.2.1 Estrutura de dados

De forma a complementar o enunciado na secção 4.1, foi criada uma estrutura principal denominada *Music*, onde se guarda a informação geral da música temporariamente até esta ser imprimida para um ficheiro.

Nesta estrutura ir-se-à guardar o titulo, o nome do autor e a letra da música, entre outros campos do cabeçalho que possam ser necessários.

A letra da musica é guardada numa lista ligada onde cada nodo é um linha da letra e é representada pela estrutura `MusicLine`. A estrutura pode ser encontrada em anexo [6.3.2](#).

### 4.2.2 Filtro de Texto

Para a filtragem do texto foram criadas várias expressões regulares, sendo que o ficheiro pode ser encontrado em anexo [\(6.3.1\)](#).

As primeiras expressões regulares, do tipo `^title: .+` servem para capturar os cabeçalhos que serão necessários. Existem também as expressões regulares do tipo: `from`, `author`, `lyrics`, `music` e `singer`, todas equivalentes. De forma a ignorar qualquer outro campo do cabeçalho que não tivesse sido previsto foi ainda criada a seguinte expressão regular: `^[a-zA-Z]+: .+`.

Quanto à deteção da letra da música existem duas expressões regulares: uma para capturar uma linha da letra, outra para capturar as linhas em branco entre os poemas, que são respetivamente: `[ ].*` e `$^\\n`.

Tal como foi dito na análise [\(4.1\)](#), existem algumas anotações no meio da letra da música que são necessárias retirar. Para isso foram criadas as seguintes expressões regulares:

- `{abc}(.|\\n)*{abcclose}` para retirar a pauta da musica.
- `[ ].*` que ignora as notas no meio dos poemas (pois estas tem um espaço no inicio).

Ainda assim estas duas expressões regulares não eram suficientes e na deteção de uma linha da letra, antes de guardar a linha, processa-se a linha com o auxílio de duas funções: `takeOffAnnotations` e `takeOffUnderScore`. A primeira retira anotações que estejam na mesma linha, e a segunda tira os caracteres '`_`' que estejam no meio da linha.

### 4.2.3 Funcionamento

De forma a perceber melhor o funcionamento do autómato esta secção irá detalhar a ponte entre o filtro de texto [\(4.2.2\)](#) e a estrutura de dados [\(4.2.1\)](#).

À medida que o autómato captura os campos do cabeçalho da música, guarda a informação com as funções de `append`, p.e. `appendAuthor`, `appendLyrics`, entre outras. Estas funções guardam os campos na variável `Music`.

As linhas da letra são guardadas através das funções `appendLine` e `appendWhiteLine`.

Quando é detetado o início de uma nova música, através de expressão regular, é executado `commitCheckNext()` que escreve a letra que está atualmente na variável `Music` para o ficheiro `latex`. Caso seja detetada a falta de algum item obrigatório então a escrita para o ficheiro é cancelada. De seguida a variável `Music` é reiniciada para a música seguinte com a função `Start()`.

Na escrita do ficheiro `latex` a letra é escrita entre as `tags` da `latex` de `Verbatim` para evitar erros no `latex` por falta de caracteres escape.

## 4.3 Testes realizados

Estão documentados neste secção 3 testes realizados ao autómato, utilizando como input os ficheiros que se encontram em anexo: [6.3.4](#), [6.3.4](#) e [6.3.4](#).

### Teste nº 1

Após a utilização do autómato no ficheiro [6.3.4](#), este gerou o output ([6.3.4](#)). Este ficheiro não tem nenhuma situação excepcional, é um caso normal.

### Teste nº 2

Após a utilização do autómato no ficheiro [6.3.4](#), este gerou o output ([6.3.4](#)). Este ficheiro tem duas situações excepcionais, o carácter '\_' no meio de palavras e notas musicais no fim das frases. Podemos verificar no output que apenas tem a letra da musica.

### Teste nº 3

Após a utilização do autómato no ficheiro [6.3.4](#), este gerou o output ([6.3.4](#)). Este ficheiro tem uma situação excepcional, antes da letra da musica tem as tags <abc>...</abc> com anotações de notas musicas. Podemos verificar que no output já não está presente.

Um exemplo de uma possível utilização do autómato é:

```
cat <inputFile> | ./play <output1.tex> <output2.tex> ...
```

## Capítulo 5

# Conclusão

Terminado o desenvolvimento do trabalho, é importante referir que o mesmo nos permitiu aprofundar o conhecimento acerca do Gerador Léxico Flex assim como da análise léxica no geral, obrigando-nos também a utilizar ferramentas tais como HTML e Latex.

Relativamente ao problema do "Museu da Pessoa", a dificuldade recaiu na definição da estrutura de suporte de dados, dado que devido à falta de claridade do enunciado foi necessário re-implementar a estrutura de forma a que esta admitisse a funcionalidade de ter um índice geral em HTML.

No problema de "Processamento de Entidades Nomeadas" foi necessário chegar a um consenso acerca das tags que deveriam ser validadas e o relacionamento possível que estas teriam entre si. Após essa decisão, o desenvolvimento da estrutura que suporta esta informação tornou-se relativamente simples.

O problema "Processamento de Ficheiros com Canções" foi resolvida através da implementação de uma estrutura capaz de evitar que a ordem dos dados no ficheiro não seja significativa (e então armazena em memória a informação). A presença de certos elementos em datasets mais diversos (tais como header's não esperados e anotações em músicas) deram origem a problemas, sendo que a solução consistiu em ignorar essa informação.

Cada elemento do grupo realizou um exercício do enunciado proposto, apoiando-se mutuamente na existência de dificuldades. Apesar das dificuldades iniciais, encontramos-nos satisfeitos com o resultado final e estamos confiantes para o próximo trabalho.

# Capítulo 6

## Anexos

### 6.1 Museu da Pessoa — tratamento de fotografias

#### 6.1.1 Filtro de Texto

```
%{
#include "album.h"
char* token;
photo_ptr picture;
%}

%%
\\<foto[ \\t]+[a-zA-Z]+=\\\".*\\\" {
yytext[strlen(yytext)-1]=\\'\\0';
picture=initPhoto(strchr(yytext,'\\')\\'+1);
}
\\<quando[ \\t]+[a-zA-Z]+=\\\"[0-9]{4}(.|-)\\{0-9}{2}(.|-)\\{0-9}{2} {
setDate(picture,strchr(yytext,'\\')\\'+1);
}
\\<quem>[ \\t\\na-zA-\\u,\\\"0-9;:\\]+ {
token=strchr(yytext,'\\')\\'+1;
setWho(picture,token);
}
\\<onde>[ \\t\\n0-9a-zA-\\u.,;:\\\"]+<{
yytext[strlen(yytext)-1]=\\'\\0';
setLoc(picture,strchr(yytext,'\\')\\'+1);
}
\\<facto>[ \\t\\n0-9a-zA-\\u.,;:\\\"]+<{
yytext[strlen(yytext)-1]=\\'\\0';
setFact(picture,strchr(yytext,'\\')\\'+1);
}
\\</foto>{
```

```

if(picture) {
    token=strdup(picture->who);
    token=strtok(token,";");
    while (token){
        insert_photo_person(token, picture);
        token=strtok(NULL,";");
    }
}
picture=NULL;
}
.|\\n ;

```

%%

### 6.1.2 Estrutura de dados

```

#include <stdio.h>
#include <string.h>
#include <stdlib.h>
#include "album.h"

static Album album;

char* trimspace(char *str)
{
    char *end;

    while (isspace(*str) || (*str)=='\n') str++;

    if (*str == 0)
        return str;

    end = str + strlen(str) - 1;
    while (end > str && isspace(*end)) end--;

    *(end+1) = 0;
    return str;
}

char* subsPoints(char *str){
    int i;

    for (i=0; i<strlen(str); i++){
        if (str[i]=='.')
            str[i]='-';
    }
    return str;
}

```

```

}

void insert_photo(photo_ptr p, photo_ptr *pics){
    if ((*pics) == NULL)
    {
        (*pics)=p;
        (*pics)->left=NULL;
        (*pics)->right=NULL;
    }
    else{
        int comp=strcmp(p->date , (*pics)->date );
        if(comp < 0)
        {
            insert_photo(p , &(*pics)->left );
        }
        else {
            insert_photo(p , &(*pics)->right );
        }
    }
}

void insert_photo_person(char *name, photo_ptr p){

    person_ptr aux, newPers, lastPers;
    int done=0, comp;

    name=trimspace(name);

    if (album.people == NULL)
    {
        album.people=malloc(sizeof(person_node));
        name=trimspace(name);
        album.people->name=strdup(name);
        album.people->href=malloc(16*sizeof(char));
        sprintf(album.people->href , "%d" ,album.count );
        album.people->href=strcat(album.people->href , ".html");
        album.people->pics=p;
        album.count++;
        album.people->next=NULL;
    }
    else{
        for (aux=album.people ; aux && (comp=strcmp (aux->name , name))<0;aux=aux->next ){
            lastPers=aux;
        }
        if (comp==0){
            insert_photo(p,&aux->pics );
        }
    }
}

```

```

        else {
            newPers = malloc(sizeof(person_node));
            newPers->name=strdup(name);
            newPers->href=malloc(16*sizeof(char));
            sprintf(newPers->href,"%d",album.count++);
            newPers->href=strcat(newPers->href,".html");
            newPers->pics=p;
            if(!aux) {
                newPers->next=NULL;
                if(lastPers) lastPers->next=newPers;
            }
            else {
                newPers->next=aux;
                if (lastPers) lastPers->next=newPers;
                else album.people=newPers;
            }
        }
    }

void setDate(photo_ptr p, char *photoD){
    photoD=trimSpace(photoD);
    photoD=subsPoints(photoD);
    p->date=strdup(photoD);
}

void setLoc(photo_ptr p, char *photoL){
    photoL=trimSpace(photoL);
    p->loc=strdup(photoL);
}

void setFact(photo_ptr p, char *photoF){
    photoF=trimSpace(photoF);
    p->fact=strdup(photoF);
}

void setWho(photo_ptr p, char *photoW){
    photoW=trimSpace(photoW);
    p->who=strdup(photoW);
}

photo_ptr initPhoto(char *photoN){
    photo_ptr pic = malloc(sizeof(photo_node));
    pic->file=strdup(photoN);
    pic->date=strdup("Desconhecida");
    pic->loc=strdup("Desconhecido");
    pic->fact=strdup("Desconhecida");
    pic->who=strdup("Nao identificados");
}

```

```

    pic->left=NULL;
    pic->right=NULL;
    return pic;
}

void init(){
    album.people=NULL;
    album.count=1;
}

void appendPhoto(photo_ptr p, FILE* file){
    if(p!=NULL) {
        appendPhoto(p->left, file);
        fprintf(file, "<div class=\"photo\">\n</br>\n");
        fprintf(file, "<div class=\"fact\"><h3>Descrição:</h3>");
        fprintf(file, "<%s</div>\n</br>\n", p->fact);
        fprintf(file, "<div class=\"img\" align=\"center\">\n");
        fprintf(file, "<img src=\"%s\" alt=\"%s\"", p->file, p->file);
        fprintf(file, " width=800 height=600>\n");
        fprintf(file, "</div>\n</br>\n</br>\n");
        fprintf(file, "<div class=\"data\"><b>Data:</b> %s</div>\n", p->date);
        fprintf(file, "<div class=\"local\"><b>Local:</b> %s</div>\n", p->loc);
        fprintf(file, "<div class=\"who\"><b>Quem:</b> %s</div>\n", p->who);
        fprintf(file, "</div>\n");
        appendPhoto(p->right, file);
    }
}

void createPage(char* filename, char* name, photo_ptr p){
    FILE* file;
    photo_ptr pic_ptr;

    file=fopen(filename, "w+");

    fprintf(file, "<!DOCTYPE html>\n<html>\n");
    fprintf(file, "<head>\n<meta charset=\"UTF-8\">\n</head>\n<body>\n");

    fprintf(file, "<a href=\"AlbumGerado.html\">Voltar</a>");

    fprintf(file, "<div class=\"title\" align=\"center\">\n");
    fprintf(file, "<h1>Album de fotografias </h1>\n");
    fprintf(file, "<h2>%s</h2>\n</div>\n</br>\n</br>\n", name);

    fprintf(file, "<div class=\"gallery\">\n");
    appendPhoto(p, file);
    fprintf(file, "</div>\n</body>\n</html>");
}

```

```

        fclose(file);
    }

void createAlbum(char* filename) {
    FILE* file;
    person_ptr aux = album.people;

    file=fopen(filename , "w+");

    fprintf(file , "<!DOCTYPE html>\n<html>\n");
    fprintf(file , "<head>\n<meta charset=\"UTF-8\">\n</head>\n<body>\n");

    fprintf(file , "<div class=\"title\" align=\"center\">\n");
    fprintf(file , "<h1>Album de fotografias</h1>\n</div>\n");

    fprintf(file , "<div class=\"index\">\n");
    fprintf(file , "<div class=\"indextitle\" align=\"center\">\n");
    fprintf(file , "<h2>Indice de Pessoas</h2>\n</div>\n");
    fprintf(file , "<div class=\"indexitems\">\n<ul>\n");

    while(aux){
        fprintf(file , "<li><a href=\"%s\"%s</a></li>\n" , aux->href , aux->name);
        createPage(aux->href , aux->name, aux->pics );
        aux=aux->next ;
    }

    fprintf(file , "</ul>\n</div>\n</div>\n");
    fprintf(file , "</body>\n</html>");

    fclose(file);
}

int main(int argc , char* argv []){
    init ();
    yylex ();

    if (argc==2){
        createAlbum(argv[1]);
    }
    else{
        createAlbum ("AlbumGerado.html");
    }
    return 0;
}

```

### 6.1.3 Cabeçalho ficheiro C

```
#ifndef __album_h__
#define __album_h__

typedef struct sPhoto *photo_ptr;

typedef struct sPhoto {
    char* file;
    char* date;
    char* loc;
    char* fact;
    char* who;
    photo_ptr left;
    photo_ptr right;
} photo_node;

typedef struct person *person_ptr;

typedef struct person
{
    char* name;
    char* href;
    photo_ptr pics;
    person_ptr next;
} person_node;

typedef struct sAlbum {
    person_ptr people;
    int count;
} Album;

void setDate(photo_ptr p, char *photoD);
void setLoc(photo_ptr p, char *photoL);
void setFact(photo_ptr p, char *photoF);
void setWho(photo_ptr p, char *photoW);
photo_ptr initPhoto(char *photoN);
void insert_photo_person(char *name, photo_ptr p);

#endif
```

### 6.1.4 Testes

#### Input teste 1

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE fotos SYSTEM "../Users/sony/mp/fotos.dtd">
<fotos>
```

```

<foto ficheiro="026-F-01.jpg">
    <onde>Taberna Necá éChamin, Afurada</onde>
    <quando data="2000-09-12"/>
    <quem>Manuel de Oliveira éChamin</quem>
    <facto>Manuel de Oliveira éChamin áatrs do ábalco na taberna
        Necá éChamin.</facto>
</foto>
<foto ficheiro="026-F-02.jpg">
    <onde>Taberna Necá éChamin, Afurada</onde>
    <quando data="2000-09-12"/>
    <quem>Manuel de Oliveira éChamin</quem>
    <facto>O interior da taberna Necá éChamin situada na rua 27
        de Fevereiro nº197, Afurada.</facto>
</foto>
</fotos>

```

#### Output teste 1

```

<!DOCTYPE html>
<html>
<head>
    <meta charset="UTF-8">
</head>
<body>
<div class="title" align="center">
    <h1>Album de fotografias</h1>
</div>
<div class="index">
    <div class="indextitle" align="center">
        <h2>Indice de Pessoas</h2>
    </div>
    <div class="indexitems">
        <ul>
            <li><a href="#">Manuel de Oliveira éChamin</a></li>
        </ul>
    </div>
</div>
</body>
</html>

```

## Album de fotografias

### Indice de Pessoas

- [Manuel de Oliveira Chaminé](#)

Figura 6.1: Indice HTML gerado pelo ficheiro XML

```

<!DOCTYPE html>
<html>
<head>
    <meta charset="UTF-8">
</head>
<body>
    <a href="AlbumGerado.html">Voltar</a><div class="title" align="center">
        <h1>Album de fotografias</h1>
        <h2>Manuel de Oliveira éChamin</h2>
    </div>
    <br>
    <br>
    <div class="gallery">
        <div class="photo">
            <br>
            <div class="fact"><b>Descrição:</b> Manuel de Oliveira éChamin  

                áatrs do ábalco na taberna  

                Necá éChamin.</div>
        </br>
        <div class="img" align="center">
            
        </div>
    </br>
    <br>
    <div class="data"><b>Data:</b> 2000-09-12</div>
    <div class="local"><b>Local:</b> Taberna Necá éChamin, Afurada</div>
    <div class="who"><b>Quem:</b> Manuel de Oliveira éChamin</div>
    <div class="photo">
        <br>
        <div class="fact"><b>Descrição:</b> O interior da taberna Necá  

            éChamin situada na rua 27 de Fevereiro nº197, Afurada.</div>
    </br>
    <div class="img" align="center">
        
    </div>
    <br>
    <br>
    <div class="data"><b>Data:</b> 2000-09-12</div>
    <div class="local"><b>Local:</b> Taberna Necá éChamin, Afurada</div>
    <div class="who"><b>Quem:</b> Manuel de Oliveira éChamin</div>
    </div>
</div>

```

</html>

[Voltar](#)

## Album de fotografias

Manuel de Oliveira Chaminé

**Descrição:**

Manuel de Oliveira Chaminé atrás do balcão na taberna Necá Chaminé.



Data: 2000-09-12  
Local: Taberna Necá Chaminé, Afurada  
Quem: Manuel de Oliveira Chaminé

Figura 6.2: Pagina HTML gerada pelo ficheiro XML (1 de 2)

**Descrição:**

O interior da taberna Necá Chaminé situada na rua 27 de Fevereiro nº197, Afurada.



Data: 2000-09-12  
Local: Taberna Necá Chaminé, Afurada  
Quem: Manuel de Oliveira Chaminé

Figura 6.3: Pagina HTML gerada pelo ficheiro XML (2 de 2)

### Input teste 2

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE fotos SYSTEM "Users/sony/mp/fotos.dtd">
<fotos>
```

```

<foto ficheiro="028-F-01.jpg">
    <onde>Taberna do Fausto</onde>
    <quando data="2000-09-22"/>
    <quem>Jos óAntnio Nunes da Silva , mais conhecido por "
        Matroco"</quem>
</foto>
<foto ficheiro="030-F-01.jpg">
<onde>Taberna do Fausto</onde>
<quando data="2000.09.22"/>
<quem>úJlio Rodrigues Barbado , o pescador mais velho da Afurada.</
    quem>
</foto>
<foto ficheiro="030-F-02.jpg">
<onde>Taberna do Fausto</onde>
<quando data="2000.09.22"/>
<quem>úJlio Rodrigues Barbado com um amigo.</quem>
</foto>
<foto ficheiro="031-F-01.jpg">
<onde>Taberna do Fausto</onde>
<quando data="2000.09.22"/>
<quem>Francisco de Sousa , cujo alcunha é "Chico_êFrancs".</quem>
</foto>
<foto ficheiro="031-F-02.jpg">
<onde>Taberna do Fausto</onde>
<quando data="2000.09.22"/>
<quem>Francisco de Sousa a beber um "pescador" , que é uma mistura
    de vinho tinto com 7up.</quem>
</foto>
<foto ficheiro="031-F-03.jpg">
<onde>Taberna do Fausto</onde>
<quando data="2000.09.22"/>
<quem>Francisco de Sousa com os amigos.</quem>
</foto>
</fotos>

```

## Output teste 2

```

<!DOCTYPE html>
<html>
<head>
    <meta charset="UTF-8">
</head>
<body>
<div class="title" align="center">
    <h1>Album de fotografias</h1>
</div>
<div class="index">
    <div class="indextitle" align="center">
        <h2>Indice de Pessoas</h2>

```

```

</div>
<div class="indexitems">
  <ul>
    <li><a href="5.html">Francisco de Sousa a beber um "pescador", que é uma mistura de vinho tinto com 7up</a></li>
    <li><a href="6.html">Francisco de Sousa com os amigos</a></li>
    <li><a href="4.html">Francisco de Sousa, cujo alcunha é "Chico Francês"</a></li>
    <li><a href="1.html">éJos Antnio Nunes da Silva , mais conhecido por "Matroco"</a></li>
    <li><a href="3.html">Jlio Rodrigues Barbado com um amigo</a></li>
    <li><a href="2.html">Jlio Rodrigues Barbado, o pescador mais velho da Afurada</a></li>
  </ul>
</div>
</div>
</body>
</html>

```

## Album de fotografias

### Indice de Pessoas

- [Francisco de Sousa a beber um "pescador", que é uma mistura de vinho tinto com 7up](#)
- [Francisco de Sousa com os amigos](#)
- [Francisco de Sousa, cujo alcunha é "Chico Francês"](#)
- [José Antônio Nunes da Silva, mais conhecido por "Matroco"](#)
- [Júlio Rodrigues Barbado com um amigo](#)
- [Júlio Rodrigues Barbado, o pescador mais velho da Afurada](#)

Figura 6.4: Indice HTML gerado pelo ficheiro XML

```

<!DOCTYPE html>
<html>
<head>
  <meta charset="UTF-8">
</head>
<body>
<a href="AlbumGerado.html">Voltar</a><div class="title" align="center">
  <h1>Album de fotografias</h1>
  <h2>Francisco de Sousa a beber um "pescador", que é uma mistura de vinho tinto com 7up</h2>
</div>
<br>
<br>
<div class="gallery">
  <div class="photo">
</br>

```

```

<div class="fact"><h3>Descrição:</h3> Desconhecida</div>
</br>
<div class="img" align="center">
    
</div>
</br>
</br>
<div class="data"><b>Data:</b> 2000-09-22</div>
<div class="local"><b>Local:</b> Taberna do Fausto</div>
<div class="who"><b>Quem:</b> Francisco de Sousa a beber um "
    pescador", que é uma mistura de vinho tinto com 7up</div>
</div>
</div>
</body>
</html>

```

[Voltar](#)

### Album de fotografias

Francisco de Sousa com os amigos

**Descrição:**  
Desconhecida



**Data:** 2000-09-22  
**Local:** Taberna do Fausto  
**Quem:** Francisco de Sousa com os amigos

Figura 6.5: Pagina HTML gerada pelo ficheiro XML

```

<!DOCTYPE html>
<html>
<head>
    <meta charset="UTF-8">
</head>
<body>
<a href="AlbumGerado.html">Voltar</a><div class="title" align="center"
    >
<h1>Album de fotografias</h1>

```

```

<h2>Francisco de Sousa com os amigos</h2>
</div>
</br>
</br>
<div class="gallery">
    <div class="photo">
        <br>
            <div class="fact"><h3>Descrição:</h3> Desconhecida</div>
        <br>
            <div class="img" align="center">
                
            </div>
        <br>
    <br>
        <div class="data"><b>Data:</b> 2000-09-22</div>
        <div class="local"><b>Local:</b> Taberna do Fausto</div>
        <div class="who"><b>Quem:</b> Francisco de Sousa com os amigos</div>
    </div>
</div>
</div>
</body>
</html>

```

[Voltar](#)

### Album de fotografias

Francisco de Sousa a beber um "pescador", que é uma mistura de vinho tinto com 7up

**Descrição:**  
Desconhecida



**Data:** 2000-09-22  
**Local:** Taberna do Fausto  
**Quem:** Francisco de Sousa a beber um "pescador", que é uma mistura de vinho tinto com 7up

Figura 6.6: Pagina HTML gerada pelo ficheiro XML

### Input teste 3

```
<?xml version="1.0"?>
<!DOCTYPE fotos SYSTEM "/Users/sony/mp/fotos.dtd">
<fotos>
    <foto ficheiro="069-D-01.jpg">
        <quando data="1969-05-12"> no quartel militar do Porto</
            quando>
        <quem> Fausto Ferreira Gomes </quem>
        <facto> Uma certidão militar , foi o documento que foi
            exigido ao
            senhor Fausto Ferreira Gomes para poder tirar a carta
            de
            condução</facto>
    </foto>
    <foto ficheiro="069-F-01.jpg">
        <onde>Taberna do Fausto , na Afurada</onde>
        <quando data="2000-09-12"></quando>
        <quem>Fausto Ferreira Gomes</quem>
    </foto>
    <foto ficheiro="069-F-02.jpg">
        <onde>Taberna do Fausto , na Afurada</onde>
        <quando data="2000-09-12"></quando>
        <quem>Fausto Ferreira Gomes e um cliente</quem>
        <facto>O interior da Taberna do Fausto situada na rua Alves
            Correia
            n.º 145 , Afurada.</facto>
    </foto>
    <foto ficheiro="069-F-03.jpg">
        <onde>Monte de Santa Luzia , Viana do Castelo</onde>
        <quando>1956</quando>
        <quem>Fausto Ferreira Gomes e a esposa , Infância Silva
            Magalhães</quem>
        <facto>Passeio do casal ao Monte de Santa Luzia em Viana do
            Castelo .</facto>
    </foto>
    <foto ficheiro="069-F-04.jpg">
        <onde>Taberna do Fausto , Afurada</onde>
        <facto>Tabuleta , "Os alegres amigos do Fausto" , foi
            oferecida ao senhor
            Fausto pelos seus amigos e clientes .</facto>
    </foto><foto ficheiro="069-F-05.jpg">
        <onde>Taberna do Fausto , Afurada</onde>
        <quem> Fausto Ferreira Gomes</quem>
        <quando data="2000.09.22"></quando>
        <legenda> Fausto Ferreira Gomes a preparar um "pirolito" , uma
            mistura de vinho branco com 7up.</legenda>
    </foto>
    <foto ficheiro="069-F-06.jpg">
        <quem>Fausto Ferreira Gomes</quem>
        <quando data="2000.09.22"></quando>
```

```

<onde> Taberna Fausto , Afurada</onde>
</foto>
</fotos>
```

### Output teste 3

```

<!DOCTYPE html>
<html>
<head>
    <meta charset="UTF-8">
</head>
<body>
<div class="title" align="center">
    <h1>Album de fotografias</h1>
</div>
<div class="index">
    <div class="indextitle" align="center">
        <h2>Indice de Pessoas</h2>
    </div>
    <div class="indexitems">
        <ul>
            <li><a href="1.html">Fausto Ferreira Gomes</a></li>
            <li><a href="3.html">Fausto Ferreira Gomes e a esposa , Inf</a></li>
            <li><a href="2.html">Fausto Ferreira Gomes e um cliente</a></li>
            >
            <li><a href="4.html">Nao identificados</a></li>
        </ul>
    </div>
</div>
</body>
</html>
```

## Album de fotografias

### Indice de Pessoas

- [Fausto Ferreira Gomes](#)
- [Fausto Ferreira Gomes e a esposa, Inf](#)
- [Fausto Ferreira Gomes e um cliente](#)
- [Nao identificados](#)

Figura 6.7: Indice HTML gerado pelo ficheiro XML

```

<!DOCTYPE html>
<html>
<head>
    <meta charset="UTF-8">
</head>
<body>
```

```
<a href="AlbumGerado.html">Voltar</a><div title="align="center">
    <h1>Album de fotografias</h1>
    <h2>Fausto Ferreira Gomes e a esposa , Inf</h2>
</div>
</br>
</br>
<div class="gallery">
    <div class="photo">
</br>
    <div class="fact"><h3>Descricao:</h3> Passeio do casal ao Monte
        de Santa Luzia em Viana do
        Castelo.</div>
</br>
    <div class="img" align="center">
        
    </div>
</br>
</br>
    <div class="data"><b>Data:</b> Desconhecida</div>
    <div class="local"><b>Local:</b> Monte de Santa Luzia , Viana do
        Castelo</div>
    <div class="who"><b>Quem:</b> Fausto Ferreira Gomes e a esposa ,
        Inf</div>
    </div>
</div>
</body>
</html>
```

[Voltar](#)

## Album de fotografias

Fausto Ferreira Gomes e a esposa, Inf

**Descrição:**

Passeio do casal ao Monte de Santa Luzia em Viana do Castelo.



Data: Desconhecida  
Local: Monte de Santa Luzia, Viana do Castelo  
Quem: Fausto Ferreira Gomes e a esposa, Inf

Figura 6.8: Pagina HTML gerada pelo ficheiro XML

```
<!DOCTYPE html>
<html>
<head>
    <meta charset="UTF-8">
</head>
<body>
    <a href="AlbumGerado.html">Voltar</a><div class="title" align="center">
        <h1>Album de fotografias</h1>
        <h2>Nao identificados</h2>
    </div>
    <br>
    <br>
    <div class="gallery">
        <div class="photo">
            <br>
            <div class="fact"><h3>Descricao:</h3> Tabuleta , "Os alegres amigos do Fausto", foi oferecida ao senhor Fausto pelos seus amigos e clientes.</div>
        <br>
        <div class="img" align="center">
            
        </div>
    <br>
    <br>
    <div class="data"><b>Data:</b> Desconhecida</div>
```

```

<div class="local"><b>Local:</b> Taberna do Fausto, Afurada</div>
<div class="who"><b>Quem:</b> Nao identificados</div>
</div>
</div>
</body>
</html>

```

[Voltar](#)

### Album de fotografias

Nao identificados

**Descrição:**

Tabuleta, "Os alegres amigos do Fausto", foi oferecida ao senhor Fausto pelos seus amigos e clientes.



**Data:** Desconhecida  
**Local:** Taberna do Fausto, Afurada  
**Quem:** Nao identificados

Figura 6.9: Pagina HTML gerada pelo ficheiro XML

### Input teste 4

```

<fotos>
<foto ficheiro="022-F-01.jpg">
<quando data="1961-01-15"/>
<quem>Ana de Lourdes de Oliveira Chamine; Antonio Oliveira Machado</
    quem>
<facto> Os noivos cortam o bolo de casamento</facto>
</foto>
<foto ficheiro="022-F-02.jpg">
<onde>Casa Machado, Afurada, Vila Nova de Gaia</onde>
<quando data="2000-09-12"/>
<quem>Ana de Lourdes de Oliveira Chamine; Antonio Oliveira Machado</
    quem>
<facto>Antonio Machado e a sua esposa, dona Ana atras do balcao da
    taberna Casa Machado.</facto>
</foto>
</fotos>

```

#### **Output teste 4**

```
<!DOCTYPE html>
<html>
<head>
    <meta charset="UTF-8">
</head>
<body>
<div class="title" align="center">
    <h1>Album de fotografias</h1>
</div>
<div class="index">
    <div class="indextitle" align="center">
        <h2>Indice de Pessoas</h2>
    </div>
    <div class="indexitems">
        <ul>
            <li><a href="1.html">Ana de Lourdes de Oliveira Chamine</a></li>
            >
            <li><a href="2.html">Antonio Oliveira Machado</a></li>
        </ul>
    </div>
</div>
</body>
</html>
```

### **Album de fotografias**

#### **Indice de Pessoas**

- [Ana de Lourdes de Oliveira Chamine](#)
- [Antonio Oliveira Machado](#)

Figura 6.10: Indice HTML gerado pelo ficheiro XML

```
<!DOCTYPE html>
<html>
<head>
    <meta charset="UTF-8">
</head>
<body>
<a href="AlbumGerado.html">Voltar</a><div class="title" align="center">
    <h1>Album de fotografias</h1>
    <h2>Ana de Lourdes de Oliveira Chamine</h2>
</div>
</br>
</br>
```

```
<div class="gallery">
    <div class="photo">
</br>
    <div class="fact"><b>Descricao:</b> Os noivos cortam o bolo de
        casamento</div>
</br>
    <div class="img" align="center">
        
    </div>
</br>
</br>
    <div class="data"><b>Data:</b> 1961-01-15</div>
    <div class="local"><b>Local:</b> Desconhecido</div>
    <div class="who"><b>Quem:</b> Ana de Lourdes de Oliveira Chamine;
        Antonio Oliveira Machado</div>
    </div>
    <div class="photo">
</br>
    <div class="fact"><b>Descricao:</b> Antonio Machado e a sua
        esposa, dona Ana atras do balcao da taberna Casa Machado.</div>
    </div>
</br>
    <div class="img" align="center">
        
    </div>
</br>
</br>
    <div class="data"><b>Data:</b> 2000-09-12</div>
    <div class="local"><b>Local:</b> Casa Machado, Afurada, Vila Nova
        de Gaia</div>
    <div class="who"><b>Quem:</b> Ana de Lourdes de Oliveira Chamine;
        Antonio Oliveira Machado</div>
    </div>
</div>
</body>
</html>
```

[Voltar](#)

## Album de fotografias

Ana de Lourdes de Oliveira Chamine

**Descricao:**

Os noivos cortam o bolo de casamento



**Data:** 1961-01-15

**Local:** Desconhecido

**Quem:** Ana de Lourdes de Oliveira Chamine; Antonio Oliveira Machado

Figura 6.11: Pagina HTML gerada pelo ficheiro XML

```
<!DOCTYPE html>
<html>
<head>
    <meta charset="UTF-8">
</head>
<body>
    <a href="AlbumGerado.html">Voltar</a><div class="title" align="center">
        <h1>Album de fotografias</h1>
        <h2>Antonio Oliveira Machado</h2>
    </div>
    <br>
    <br>
    <div class="gallery">
        <div class="photo">
        </div>
        <div class="fact"><h3>Descricao:</h3> Os noivos cortam o bolo de
            casamento</div>
    </div>
    <div class="img" align="center">
        
    </div>
    <br>
    <br>
    <div class="data"><b>Data:</b> 1961-01-15</div>
    <div class="local"><b>Local:</b> Desconhecido</div>
```

```

<div class="who"><b>Quem:</b> Ana de Lourdes de Oliveira Chamine;  

    Antonio Oliveira Machado</div>
</div>
<div class="photo">
</br>
<div class="fact"><h3>Descricao:</h3> Antonio Machado e a sua  

    esposa, dona Ana atrás do balcão da taberna Casa Machado.</div>
</>
</br>
<div class="img" align="center">
    
</div>
</br>
</br>
<div class="data"><b>Data:</b> 2000-09-12</div>
<div class="local"><b>Local:</b> Casa Machado, Afurada, Vila Nova  

    de Gaia</div>
<div class="who"><b>Quem:</b> Ana de Lourdes de Oliveira Chamine;  

    Antonio Oliveira Machado</div>
</div>
</div>
</body>
</html>

```

**Descrição:**

Antonio Machado e a sua esposa, dona Ana atrás do balcão da taberna Casa Machado.



**Data:** 2000-09-12  
**Local:** Casa Machado, Afurada, Vila Nova de Gaia  
**Quem:** Ana de Lourdes de Oliveira Chamine; Antonio Oliveira Machado

Figura 6.12: Pagina HTML gerada pelo ficheiro XML

## 6.2 Processamento de Entidades Nomeadas (Enamex)

### 6.2.1 Filtro de Texto

```
%s country

%{
#include "tree.h"

char *aux;
char* aux_subcity;
int aux_len;
tree_ptr persons=NULL;
tree_ptr countrys = NULL;
tree_ptr citys = NULL;
tree_ptr organizations = NULL;
tree_ptr otherlocations = NULL;
char* trimspace(char *str);

}

pal      [a-zA-Z0-9Ç-ÑÀ-Ù ]+
enamex  \<[ \t]*(?i:enamex)[ \t]+(?i:type)=
eclose   \<[ \t]*\/(?i:enamex)[ \t]*\>
org      \\"[ \t]*(?i:organization)[ \t]*\"[ \t]*\>
person   \\"[ \t]*(?i:person)[ \t]*\"[ \t]*\>
loc      \\"[ \t]*(?i:location)[ \t]*\"[ \t]*\(>)?"
subcountry (?i:subtype)=\\"[ \t]*(?i:country)[ \t]*\"[ \t]*\>
subcity   (?i:subtype)=\\"[ \t]*(?i:city)[ \t]*\"[ \t]*\>

%%

BEGIN 0;

{enamex}{loc}{subcountry}{pal}{eclose}  {
    aux=strdup(strchr(yytext,'>')+1);
    *strchr(aux,'<')='\'0';
    insert(trimspace(aux), &countrys);
    BEGIN country;
}

<country>\.          BEGIN 0;

<country>{enamex}{loc}{subcity}{pal}{eclose}  {
    tree_ptr aux_tree = search_tree(aux, countrys);
    aux_subcity=strdup(strchr(yytext,'>')+1);
    *strchr(aux_subcity,'<')='\'0';
    insert_subtree(trimspace(aux_subcity), &countrys);
    insert(trimspace(aux_subcity), &citys);
```

```

        BEGIN 0;
    }
{enamex}{person}{pal}{eclose}  {
    aux=strdup(strchr(yytext,'>')+1);
    *strchr(aux,'<')='\\0';
    insert(trimspace(aux), &persons);
}
{enamex}{org}{pal}{eclose}  {
    aux=strdup(strchr(yytext,'>')+1);
    *strchr(aux,'<')='\\0';
    insert(trimspace(aux), &organizations);
}
{enamex}{loc}{subcity}{pal}{eclose}  {
    aux=strdup(strchr(yytext,'>')+1);
    *strchr(aux,'<')='\\0';
    insert(trimspace(aux), &citys);
}
{enamex}{loc}{pal}{eclose}  {
    aux=strdup(strchr(yytext,'>')+1);
    *strchr(aux,'<')='\\0';
    insert(trimspace(aux), &otherlocations);
}

<*>. | \\n
;

%%

/*Função que retira os espaços brancos dos nomes*/
char* trimspace(char *str)
{
    char *end;

    while(isspace(*str)) str++;

    if(*str == 0)
        return str;

    end = str + strlen(str) - 1;
    while(end > str && isspace(*end)) end--;

    *(end+1) = 0;
    return str;
}

void printHeader(FILE *f){

    fprintf(f, "<!DOCTYPE html>\\n<html>\\n<head> <meta charset=\"UTF-8\">\\n Grupo 7</head>\\n");
    fprintf(f, "<link type=\"text/css\" rel=\"stylesheet\" href=\"stylesheet.css\"/>");
}

```

```

        fprintf(f, "\t<title>Processamento de Linguagens</title><body>");

    }

void createIndex(FILE *f){
    printHeader(f);
    fprintf(f, "<h1>Processamento de Linguagens</h1></a>\n");
    fprintf(f, "<h2>Processamento de Entidades Nomeadas</h2></a>\n");
    fprintf(f, "<a href=\"persons.html\"><p class=\"ind\">Persons</p></a>\n");
    fprintf(f, "<a href=\"countrys.html\"><p class=\"ind\">Countrys</p></a>\n");
    fprintf(f, "<a href=\"organizations.html\"><p class=\"ind\">Organizations</p></a>\n");
    fprintf(f, "<a href=\"citys.html\"><p class=\"ind\">City's</p></a>\n");
    fprintf(f, "<a href=\"otherlocations.html\"><p class=\"ind\">Other Locations</p></a>\n");
    fprintf(f, "\t</body>\n</html>\n");
}

int yywrap()
{ return(1); }

int main()
{
    yylex();

    FILE *i, *p, *c, *o, *ac, *ol;

    i = fopen("web/index.html", "w");
    createIndex(i);
    fclose(i);

    p = fopen("web/persons.html", "w");
    printHeader(p);
    fprintf(p, "<h1>Persons</h1>");
    treeToHtml(persons,p);
    fprintf(p, "<a href=\"index.html\"><p>Voltar</p></a>\n\t</body>\n</html>\n");
    fclose(p);

    c = fopen("web/countrys.html", "w");
    printHeader(c);
    fprintf(c, "<h1>Countrys</h1>");
    treeToHtml(countrys,c);
    fprintf(c, "<a href=\"index.html\"><p>Voltar</p></a>\n\t</body>\n</html>\n");
    fclose(c);

    o = fopen("web/organizations.html", "w");
    printHeader(o);
    fprintf(o, "<h1>Organizations</h1>");
    treeToHtml(organizations,o);
    fprintf(o, "<a href=\"index.html\"><p>Voltar</p></a>\n\t</body>\n</html>\n");
    fclose(o);
}

```

```

ac = fopen("web/citys.html", "w");
printHeader(ac);
fprintf(ac, "<h1>City's</h1>");
treeToHtml(citys,ac);
fprintf(ac, "<a href=\"index.html\"><p>Voltar</p></a>\n\t</body>\n</html>\n");
fclose(ac);

ol = fopen("web/otherlocations.html", "w");
printHeader(ol);
fprintf(ol, "<h1>Other Locations</h1>");
treeToHtml(otherlocations,ol);
fprintf(ol, "<a href=\"index.html\"><p>Voltar</p></a>\n\t</body>\n</html>\n");
fclose(ol);

return 0;
}

```

### 6.2.2 Estrutura de dados

```

#include <stdlib.h>
#include <stdio.h>
#include <string.h>

#include "tree.h"

typedef struct tree
{
    char desig[MAX_SIZE];
    int num;
    tree_ptr left, right, subtree;
} Tree;

void insert(char *novo_desig, tree_ptr *p)
{
    if ((*p) == NULL)
    {
        (*p)=(tree_ptr) malloc(sizeof(Tree));
        strcpy((*p)->desig,novo_desig);
        (*p)->num=1;
        (*p)->left=NULL;
        (*p)->right=NULL;
        (*p)->subtree=NULL;
    }
}

```

```

else{
    int comp=strcmp(novo_desig,(*p)->desig);
    if(comp < 0)
    {
        insert(novo_desig, &(*p)->left );
    }
    else if (comp > 0)
    {
        insert(novo_desig, &(*p)->right );
    }
    else
    {
        (*p)->num++;
    }
}

void insert_subtree(char *novo_desig, tree_ptr *p){
    if ((*p) != NULL)
    {
        insert(novo_desig, &(*p)->subtree );
    }
}

void subtree_to_html(tree_ptr sub_arv, FILE *f){
    if(sub_arv != NULL){
        subtree_to_html(sub_arv->left , f );
        fprintf(f, "<ul>" );
        fprintf(f, "\t<li><p class=\"node\">%s</p></li>\n" ,sub_arv->desig );
        fprintf(f, "</ul>" );
        subtree_to_html(sub_arv->right , f );
    }
}

void treeToHtml(tree_ptr btree , FILE *f){
    if (btree != NULL)
    {
        treeToHtml(btree->left , f );
        fprintf(f, "<p class=\"root\">%s</p>\n" ,btree->desig );

        if( btree->subtree != NULL)
            subtree_to_html(btree->subtree , f );

        treeToHtml(btree->right , f );
    }
}

void imprime_inorder_tree(tree_ptr p)

```

```

{
    if (p!=NULL)
    {
        imprime_inorder_tree(p->left );
        printf ("%s->%d\n",p->desig , p->num);
        if(p->subtree != NULL) imprime_inorder_tree(p->subtree );
        imprime_inorder_tree(p->right );
    }
}

tree_ptr copyTree (tree_ptr p)
{
    tree_ptr aux=NULL;
    if (p!=NULL) {
        insert(p->desig ,&aux );
        aux->left =copyTree(p->left );
        aux->right =copyTree(p->right );
    }
    return aux;
}

void imprime_mais_freq(tree_ptr p, int n)
{
    if (p!=NULL)
    {
        imprime_mais_freq(p->left ,n);
        if ((p->num)>n) printf ("%s\n",p->desig );
        imprime_mais_freq(p->right ,n);
    }
}

int total_Pubs(tree_ptr p)
{
    if (!p) return 0;
    else return (p->num)+total_Pubs(p->left )+total_Pubs(p->right );
}

int total_elems(tree_ptr p) {
    if (!p) return 0;
    else return 1 + total_elems(p->left )+total_elems(p->right );
}

tree_ptr search_tree(char *desig , tree_ptr p)
{

```

```

if (!p) return NULL;
else {
    int comp=strcmp(p->desig , desig );
    if (comp==0) return (p);
    else {
        if (comp>0) return search_tree(desig , p->left );
        else return search_tree(desig , p->right );
    }
}
}

void imprime_pos_order_by_N(tree_ptr p, int *N) {
    if (p){
        imprime_pos_order_by_N(p->right ,N);
        if (*N){
            printf ("Elemento: %s\n" , p->desig );
            printf ("Nmr: %d\n\n" , p->num );
            (*N)--;
        }
        imprime_pos_order_by_N(p->left ,N);
    }
}

void makeempty(tree_ptr p)
{
    if (p != NULL)
    {
        makeempty(p->left );
        makeempty(p->right );
        free(p);
    }
}

```

### 6.2.3 Cabeçalho ficheiro C

```

#ifndef ____TREE_H_____
#define ____TREE_H_____

#define MAX_SIZE 200

typedef struct tree *tree_ptr;

/*Insere um novo elemento na BS-tree*/
void insert(char *novo_desig , tree_ptr *p);

/*Imprime todos os nodos(elementos) da BS-tree por ordem alfabetica*/
void imprime_inorder_tree(tree_ptr p);

```

```

/*Imprime os elementos que aparecem no minimo n vezes*/
void imprime_mais_freq(tree_ptr p, int n);

/*Devolve numero de ocorrencias de todos os elementos da BS-tree*/
int total_Pubs(tree_ptr p);

/*Devolve numero de nodos(elementos) da BS-tree*/
int total_elems(tree_ptr p);

/*Copia uma BS-tree*/
tree_ptr copyTree (tree_ptr p);

/*Procura um elemento numa BS-tree e devolve um apontador ele caso se encontrar.*
tree_ptr search_tree(char *desig, tree_ptr p);

/*Imprime por ordem decrescente de numero de ocorrencias.*
void imprime_pos_order_by_N(tree_ptr p, int *N);

/*Liberta memoria de BS-tree*/
void makeempty(tree_ptr p);

/*Insere elementos em subarvore*/
void insert_subtree(char *novo_desig, tree_ptr *p);

/*Imprime para ficheiro uma dada subarvore*/
void treeToHtml(tree_ptr btree, FILE *f);

void subtree_to_html(tree_ptr sub_arv, FILE *f);

#endif

```

#### 6.2.4 Testes

##### Input teste 1

O meu nome é <ENAMEX TYPE="PERSON">Paulo úArajo</ENAMEX> e sou de <ENAMEX TYPE="LOCATION" SUBTYPE="COUNTRY">Portugal</ENAMEX>, que tem cidades à bonitas como <ENAMEX TYPE="LOCATION" SUBTYPE="CITY">Braga</ENAMEX>, <ENAMEX TYPE="LOCATION" SUBTYPE="CITY">Viana do Castelo</ENAMEX> e <ENAMEX TYPE="LOCATION" SUBTYPE="CITY">Lisboa</ENAMEX>. ãNo tem cidades menos bonitas como <ENAMEX TYPE="LOCATION" SUBTYPE="CITY">Texas</ENAMEX> e <ENAMEX TYPE="LOCATION" SUBTYPE="CITY">Novo Mexico</ENAMEX>.

## Output teste 1

Grupo 7

Portugal

- Braga
- Lisboa
- Viana do Castelo

[Voltar](#)

## Input teste 2

<ENAMEX TYPE="PERSON"> Bento de Castro Abreu </ENAMEX>, que emigrou para o <ENAMEX TYPE="LOCATION" SUBTYPE="CITY"> Rio de Janeiro </ENAMEX> em <TIMEX TYPE="DATE"> 1/2/1895 </TIMEX>, com 26 anos de idade, é referido no seu passaporte como áproprietrio, sobrinho de outro <ENAMEX TYPE="?"> Brasileiro </ENAMEX> <ENAMEX TYPE="PERSON"> Fernando de Castro Abreu e Magalhes </ENAMEX>, que apoiou financeiramente a çäconstruo da casa do <ENAMEX TYPE="LOCATION"> Santo Novo </ENAMEX>.

<ENAMEX TYPE="?">Foi Visconde e êMarqus de àParanagu</ENAMEX> <ENAMEX TYPE="PERSON">Francisco Vilela Barbosa</ENAMEX> que nasceu no <ENAMEX TYPE="LOCATION" SUBTYPE="COUNTRY">Brasil</ENAMEX>, em <ENAMEX TYPE="LOCATION" SUBTYPE="CITY">Parati</ENAMEX> em <ENAMEX TYPE="?">Novembro</ENAMEX> de <NUMEX>1769</NUMEX> e ali morreu em <TIMEX TYPE="DATE">Setembro de 1846</TIMEX>, filho de <ENAMEX TYPE="PERSON">Francisco Vilela Barbosa</ENAMEX>, natural de <ENAMEX TYPE="LOCATION" SUBTYPE="CITY">Braga</ENAMEX> (<ENAMEX TYPE="LOCATION" SUBTYPE="Country">Portugal</ENAMEX>) e de <ENAMEX TYPE="PERSON">D. Ana Maria da çäConceio</ENAMEX>.

Formado em ámatemtica pela <ENAMEX TYPE="ORGANIZATION">Universidade de Coimbra</ENAMEX>, em <Numex>1789</NumEX> assentou çpraa na armada e quando °2 tenente prestou relevantes çservios no cerco de <ENAMEX TYPE="LOCATION" SUBTYPE="CITY">Tunis</ENAMEX> e na çäperseguiuo aos pirata argelinos do <ENAMEX TYPE="?">âMediterrneo</ENAMEX>.

Nomeado lente da <ENAMEX TYPE="ORGANIZATION">Academia da Marinha</ENAMEX>, passou em <Numex>1801</NuMEX> para o <ENAMEX TYPE="ORGANIZATION">Real Corpo de Engenheiros</ENAMEX> e em <Numex>1810</NuMEX> foi promovido a <ENAMEX TYPE="?">Major</ENAMEX> e reformado mais tarde no posto de <ENAMEX TYPE="?">Brigadeiro</ENAMEX>.

#### Output teste 2



#### Input teste 3

```
<DOC>
<DOCID> nyt960214.0704 </DOCID>
<STORYID> cat=f pri=u A4479 </STORYID>
<SLUG> fv=t_af-z> BC-<ENAMEX TYPE="PERSON">MURDOCH</ENAMEX>-SATELLITE-
NYT </SLUG>
<DATE> <TIMEX TYPE="DATE">02-14</TIMEX> </DATE>
<NWORDS> 0608 </NWORDS>
<PREAMBLE>
BC-<ENAMEX TYPE="PERSON">MURDOCH</ENAMEX>-SATELLITE-NYT
<ENAMEX TYPE="PERSON">MURDOCH</ENAMEX> SATELLITE FOR LATIN
PROGRAMMING EXPLODES ON TAKEOFF
(kd)
By <ENAMEX TYPE="PERSON">MARK LANDLER</ENAMEX>
c.<TIMEX TYPE="DATE">1996</TIMEX> <ENAMEX TYPE="ORGANIZATION">N.Y.
Times News Service</ENAMEX>
```

</PREAMBLE>

<TEXT>

<p>

A Chinese rocket carrying a television satellite exploded seconds after launch <TIMEX TYPE="DATE">Wednesday</TIMEX>, dealing a potential blow to <ENAMEX TYPE="PERSON">Rupert Murdoch</ENAMEX>'s ambitions to offer satellite programming in <ENAMEX\_TYPE="LOCATION">Latin America</ENAMEX>.

<p>

<ENAMEX\_TYPE="PERSON">Murdoch</ENAMEX>'s <ENAMEX\_TYPE="ORGANIZATION"> News Corp.</ENAMEX> is one of four media companies in a partnership that had leased space on the <ENAMEX\_TYPE="ORGANIZATION"> Intelsat</ENAMEX> satellite to offer the Latin American service. The other partners are <ENAMEX\_TYPE="ORGANIZATION">Tele-Communications Inc.</ENAMEX>, the nation's largest cable operator; <ENAMEX\_TYPE="ORGANIZATION">Grupo Televisa SA</ENAMEX>, the Mexican broadcaster and publisher, and the giant Brazilian media conglomerate <ENAMEX\_TYPE="ORGANIZATION">Globo </ENAMEX>.

<p>

<ENAMEX\_TYPE="PERSON">Lennel Evangelista</ENAMEX>, a spokesman for <ENAMEX\_TYPE="ORGANIZATION">Intelsat</ENAMEX>, a global satellite consortium based in <ENAMEX\_TYPE="LOCATION">Washington</ENAMEX>, said the accident occurred at <TIME\_TYPE="TIME">2 p.m. EST</TIME> <TIME\_TYPE="DATE"> Wednesday</TIME>, or <TIME\_TYPE="TIME">early Thursday morning</TIME> at the <ENAMEX\_TYPE="LOCATION"> Xichang</ENAMEX> launch site in <ENAMEX\_TYPE="LOCATION"> Sichuan Province</ENAMEX> in southwestern <ENAMEX\_TYPE="LOCATION"> China</ENAMEX>. "We have no details on what caused the accident," he said.

<p>

<ENAMEX\_TYPE="PERSON">Evangelista</ENAMEX> said the Chinese-built Long March rocket veered off course and was destroyed after it failed to reach orbit. <ENAMEX\_TYPE="ORGANIZATION">Intelsat</ENAMEX> was using the Long March rocket for the first time to launch one of its satellites. <ENAMEX\_TYPE="ORGANIZATION">Intelsat</ENAMEX> currently has 23 satellites in orbit.

<p>

A spokesman for <ENAMEX\_TYPE="ORGANIZATION"> News Corp.</ENAMEX>, <ENAMEX\_TYPE="PERSON">Howard Rubenstein</ENAMEX>, said the accident would not hinder the group's plans to offer 150 channels of entertainment, news and sports programming to viewers in <ENAMEX\_TYPE="LOCATION"> Latin America</ENAMEX> and the <ENAMEX\_TYPE="LOCATION"> Caribbean</ENAMEX>.

<p>

“<ENAMEX TYPE="ORGANIZATION">News Corp.</ENAMEX> has a number of other real options and will disclose them shortly , ” <ENAMEX TYPE="PERSON">Rubinstein</ENAMEX> said in a statement .

<p><ENAMEX TYPE="ORGANIZATION">Grupo Televisa</ENAMEX> and <ENAMEX TYPE="ORGANIZATION">Globo</ENAMEX> plan to offer national and local programming in Spanish and Portuguese. Initially , the venture's partners said they planned to invest <NUMEX\_TYPE="MONEY">\$500 million </NUMEX>.

<p>But a similar explosion <TIMEX\_TYPE="DATE">last year</TIMEX> delayed the plans of several American media companies to offer a package of satellite television services in <ENAMEX\_TYPE="LOCATION">Asia</ENAMEX> . <ENAMEX\_TYPE="ORGANIZATION">Viacom</ENAMEX> , <ENAMEX\_TYPE="ORGANIZATION">Time Warner</ENAMEX> 's <ENAMEX\_TYPE="ORGANIZATION">Home Box Office</ENAMEX> and <ENAMEX\_TYPE="ORGANIZATION">Turner Broadcasting System</ENAMEX> were among the companies that had leased space on an Apstar 2 satellite to beam MTV, CNN and other channels throughout <ENAMEX\_TYPE="LOCATION">Asia</ENAMEX> .

<p>After the rocket carrying that satellite exploded , media analysts said the companies had to settle for space on a series of regional satellites , which had less reach than the Apstar 2 would have offered .

<p><ENAMEX TYPE="ORGANIZATION">News Corp.</ENAMEX> actually benefited from that accident . In <TIMEX\_TYPE="DATE">1993</TIMEX> , the company had purchased a controlling stake in a rival Asian satellite service , <ENAMEX\_TYPE="ORGANIZATION" STATUS="OPT">Star TV</ENAMEX>. With his biggest competitors unable to enter the Asian market , <ENAMEX\_TYPE="PERSON">Murdoch</ENAMEX> was able to build <ENAMEX\_TYPE="ORGANIZATION" STATUS="OPT">Star TV</ENAMEX> into the dominant programming service .

<p>A spokeswoman for <ENAMEX\_TYPE="ORGANIZATION">Tele-Communications</ENAMEX> , <ENAMEX\_TYPE="PERSON">LaRae Marsik</ENAMEX> , said the partners in the Latin American venture intended to begin service by <TIMEX\_TYPE="DATE">the end of 1996</TIMEX> . When the companies announced their plans <TIMEX\_TYPE="DATE">last November</TIMEX> , they said they planned to be in business by <TIMEX\_TYPE="DATE">May</TIMEX> .

<p>Ms. <ENAMEX\_TYPE="PERSON">Marsik</ENAMEX> said <ENAMEX\_TYPE="ORGANIZATION">Tele-Communications</ENAMEX> and its partners had a back-up plan , which could include leasing space on another

satellite , but she declined to offer details. "It is an unfortunate incident , " she said , "but it is not a make-it -or-break-it event for us . "

<p>

<ENAMEX TYPE="PERSON">Jessica Reif</ENAMEX> , a media analyst at <ENAMEX TYPE="ORGANIZATION">Merrill Lynch & Co.</ENAMEX> , said , " If they can get up and running with exclusive programming <TIMEX TYPE="DATE" STATUS="OPT">within six months</TIMEX> , it doesn't set the venture back that far . "

<p>

<ENAMEX\_TYPE="ORGANIZATION">Hughes Electronics </ENAMEX> , a subsidiary of the <ENAMEX\_TYPE="ORGANIZATION">General Motors Corp.</ENAMEX> , is starting its own satellite broadcast service in <ENAMEX\_TYPE="LOCATION">Latin America</ENAMEX> . Ms. <ENAMEX\_TYPE="PERSON">Reif </ENAMEX> , said that venture , which is based on <ENAMEX\_TYPE="ORGANIZATION">Hughes</ENAMEX> 's <ENAMEX\_TYPE="ORGANIZATION" STATUS="OPT">DirecTV</ENAMEX> service in the <ENAMEX\_TYPE="LOCATION">United States</ENAMEX> , would benefit if the explosion delayed the <ENAMEX\_TYPE="PERSON">Murdoch</ENAMEX>-led venture .

</TEXT>

<TRAILER>

NYT<TIMEX TYPE="DATE">02-14-96</TIMEX> <TIMEX TYPE="TIME">2029EST</TIMEX>

</TRAILER>

</DOC>

### Output teste 3

Grupo 7

Globo

Grupo Televisa

Grupo Televisa SA

Home Box Office

Hughes

Hughes Electronics

Intelsat

Time Warner

Viacom

[Voltar](#)

## 6.3 Processamento de ficheiros com Canções

### 6.3.1 Filtro de Texto

```
%{
#include "musica.h"

char* takeOffAnnotations(char* c){
    int i=0;
    while(c[i] != '\0' && !(c[i] == ' ' && c[i + 1] == ' ')) {
        i++;
    }
    c[i] = '\0';
    return c;
}
```

```

char* takeOffUnderScore(char* c){
    int i=0, w=0;
    while(c[i] != '\0') {
        //ignore if '_'
        if(c[i] != '_') {
            c[w] = c[i];
            w++;
        }
        i++;
    }
    c[w] = '\0';
    return c;
}

abc      \<[ ]*(?i:abc)[ ]*\>
abcclose \<\/[ ]*(?i:abc)[ ]*\>

%%
^title:+      appendTitle(yytext + 6);
^from:+       appendFrom(yytext + 5);
^author:+     appendAuthor(yytext + 7);
^lyrics:+     appendLyrics(yytext + 7);
^music:+      appendMusic(yytext + 6);
^singer:+     appendSinger(yytext + 7);

^[a-zA-Z]+:+ ;// Ignore other headers:
{abc}(.|\n)*{abcclose} ;

##-+ {commitCheckNext(); start();}// Music End

^\n      appendWhiteLine(); // white music line.
[ ].* ; //ignore lines that start with space (they are music annotations);
.*   {appendLine(takeOffAnnotations(takeOffUnderScore(yytext)));
      //poem line, taking of underscore and music annotations}
.|\n  ;
%%
```

### 6.3.2 Estrutura de dados

```
#include <stdio.h>
#include <unistd.h>
#include "musica.h"
#include <stdlib.h>
```

```

#include <string.h>

// ##### private header area #####
// #####
// #####
extern int yylex();

typedef struct sMusicLine {
    char* line;
    struct sMusicLine* next;
} MusicLine;

typedef struct sMusic {
    char* _Title;
    char* _From;
    char* _Author;
    char* _Lyrics;
    char* _Music;
    char* _Singer;

    MusicLine* poem;
    MusicLine* poemEnd;
    int error;
} Music;

void writeLatex();

// ##### Implementation Area #####
// #####
// #####
static Music music;
static char** argv;
static int argc;
static int argIndex;

#define append(name) \
void append##name(char* t) { \
    if (music._##name != NULL) { \
        music.error = 1; \
    } else { \
        music._##name = strdup(t); \
    } \
}

append>Title)
append(From)

```

```

append (Author)
append (Lyrics)
append (Music)
append (Singer)

void appendLine(char* line) {
    if (music . poemEnd == NULL){
        music . poem = (MusicLine*) malloc (sizeof(MusicLine));
        music . poemEnd = music . poem;
    } else {
        music . poemEnd->next = (MusicLine*) malloc (sizeof(MusicLine));
        music . poemEnd = music . poemEnd->next ;
    }
    music . poemEnd->next = NULL;
    music . poemEnd->line = strdup (line);
}

void appendWhiteLine() {
    // don 't append white line in the init of the file

    if (music . poemEnd != NULL){
        music . poemEnd->next = (MusicLine*) malloc (sizeof(MusicLine));
        music . poemEnd = music . poemEnd->next ;
        music . poemEnd->next = NULL;
        music . poemEnd->line = strdup ("");
    }
}

void start () {
    music . _Title = NULL;
    music . _From = NULL;
    music . _Author = NULL;
    music . _Lyrics = NULL;
    music . _Music = NULL;
    music . _Singer = NULL;

    music . poem = NULL;
    music . poemEnd = NULL;
    music . error = 0;
}

void writeLatex (Music m, FILE* f) {
    fprintf (f, "\\title{\%s}\n",m. _Title );
    if (m. _Author != NULL) {
        fprintf (f, "\\author{\%s}\n",m. _Author );
    } else {
        fprintf (f, "\\author{\%s,\%s}\n",m. _Lyrics ,m. _Music );
    }
}

```

```

fprintf(f, "\\documentclass[12 pt]{ article}\\n");
fprintf(f, "\\begin{document}\\n");
fprintf(f, "\\maketitle\\n");
fprintf(f, "\\section*{Letra}\\n");
fprintf(f, "\\begin{center}\\n");
MusicLine* it;
fprintf(f, "\\begin{verbatim}\\n");
for (it = m.poem; it ; it = it->next){
    // insert line
    //if (it->line[0] != '| 0'){
        fprintf(f, "%s\\n", it->line);
    //}
    else {
        // insert white space
        //fprintf(f, "||vspace{5mm}|n");
    }
}
fprintf(f, "\\end{verbatim}\\n");
if (m._Singer){
    fprintf(f, "\\vspace{5mm}\\n");
    fprintf(f, "\\hfill %s\\n", m._Singer);
}
fprintf(f, "\\end{center}\\n");
fprintf(f, "\\end{document}\\n");
}

int commitCheckNext () {
    if (music.error != 0) {
        // need something here?
    } else {
        FILE* f;
        if (argIndex < argc) {
            f= fopen(argv[argIndex], "w");
            writeLatex(music, f);
            fflush(f);
            fclose(f);
            // execl("/usr/bin/pdflatex", "pdflatex", argv[argIndex], NULL);
        } else {
            char buffer[20];
            sprintf(buffer, "%d.tex", argIndex);
            f = fopen(buffer, "w");
            writeLatex(music, f);
            fflush(f);
            fclose(f);
            // execl("/usr/bin/pdflatex", "pdflatex", buffer, NULL);
        }
        argIndex++;
    }
}

```

```

    return music.error;
}

int main(int _argc, char* _argv[]){
    argc = _argc - 1; // first argument is the program name.
    argv = (char**) malloc(sizeof(char) * _argc);
    for (_argc--; _argc > 0; _argc--) {
        argv[_argc - 1] = strdup(_argv[_argc]);
    }
    start();
    yylex();
    commitCheckNext();
}

int yywrap() {
    return -1;
}

```

### 6.3.3 Cabeçalho ficheiro C

```

#ifndef __musica_h__
#define __musica_h__

void appendTitle(char* title);
void appendFrom(char* from);
void appendAuthor(char* author);
void appendLine(char* line);
void appendWhiteLine();
void start();
int commitCheckNext();

#endif

```

### 6.3.4 Testes

#### Input teste 1

```

title: Amêndoа Amarga
lyrics: José Carlos Ary dos Santos
music: Alain Oulman
singer: Amália Rodrigues

Port ti falo
e ninguém pensa
mas eu digo

```

minha amêndoaa, meu amigo  
meu irmão  
meu tropel de ternura  
minha casa  
meu jardim de carência  
minha asa.

Por ti vivo  
e ninguém pensa  
mas eu sigo  
um caminho de silvas  
e de nardos  
uma intensa ternura  
que persigo  
rodeada de cardos  
por tantos lados.

Por ti morro  
e ninguém sabe  
mas eu espero  
o teu corpo que sabe  
a madrugada  
o teu corpo que sabe  
a desespero

ó minha amarga amêndoaa  
desejada.

ó minha amarga amêndoaa  
desejada.

### Output teste 1

```
\title{ Amêndoaa Amarga}
\author{ José Carlos Ary dos Santos, Alain Oulman }
\documentclass[12pt]{article}
\begin{document}
\maketitle
\section*{Letra}
\begin{center}
\begin{verbatim}
Port ti falo
e ninguém pensa
mas eu digo
minha amêndoaa, meu amigo
meu irmão
meu tropel de ternura
minha casa

```

meu jardim de carência  
minha asa.

Por ti vivo  
e ninguém pensa  
mas eu sigo  
um caminho de silvas  
e de nardos  
uma intensa ternura  
que persigo  
rodeada de cardos  
por tantos lados.

Por ti morro  
e ninguém sabe  
mas eu espero  
o teu corpo que sabe  
a madrugada  
o teu corpo que sabe  
a desespero

ó minha amarga amêndoaa  
desejada.

ó minha amarga amêndoaa  
desejada.

\end{verbatim}  
\vspace{5mm}  
\hfill Amália Rodrigues  
\end{center}  
\end{document}

## Input teste 2

title: Amêndoaa Amarga  
lyrics: José Carlos Ary dos Santos  
music: Alain Oulman  
singer: Amália Rodrigues

Port ti falo  
e ninguém pensa  
mas eu digo  
minha amêndoaa, meu amigo  
meu irmão  
meu tropel de ternura  
minha casa  
meu jardim de carência  
minha asa.

Por ti vivo  
e ninguém pensa  
mas eu sigo  
um caminho de silvas  
e de nardos  
uma intensa ternura  
que persigo  
rodeada de cardos  
por tantos lados.

Por ti morro  
e ninguém sabe  
mas eu espero  
o teu corpo que sabe  
a madrugada  
o teu corpo que sabe  
a desespero

ó minha amarga amêndoaa  
desejada.

ó minha amarga amêndoaa  
desejada.

## Output teste 2

```
\title{ * Tejo que levas as águas}
\author{ Manuel da Fonseca, Adriano Correia de Oliveira}
\documentclass[12pt]{article}
\begin{document}
\maketitle
\section*{Letra}
\begin{center}
\begin{verbatim}
Tejo que levas as águas
correndo de par em par
lava a cidade de mágoas
leva as mágoas para o mar

Lava-a de crimes espantos
de roubos, fomes, terrores,
lava a cidade de quantos
do ódio fingem amores

Leva nas águas as grades
de aço e silêncio forjadas
deixa soltar-se a verdade
\end{verbatim}


```

das bocas amordaçadas

Lava bancos e empresas  
dos comedores de dinheiro  
que dos salários de tristeza  
arrecadam lucro inteiro

Lava palácios vivendas  
casebres bairros da lata  
leva negócios e rendas  
que a uns farta e a outros mata

Tejo que levas as águas  
correndo de par em par  
lava a cidade de mágoas  
leva as mágoas para o mar

Lava avenidas de vícios  
vielas de amores venais  
lava albergues e hospícios  
cadeias e hospitais

Afoga empenhos favores  
vãs glórias, ocas palmas  
leva o poder dos senhores  
que compram corpos e almas

Leva nas águas as grades  
...

Das camas de amor comprado  
desata abraços de lodo  
rostos corpos destroçados  
lava-os com sal e iodo

Tejo que levas nas águas  
...  
\end{verbatim}  
\vspace{5mm}  
\hfill Adriano Correia de Oliveira  
\end{center}  
\end{document}

### Input teste 3

```
title: = Raúl tinha um Ioio
singer: Bando dos Gambozinos
lyrics: Manuel António Pina
```

music: Suzana Ralha  
in: "o beco dos gambozinos"  
from: jj

<abc>  
X: 1  
M: 2/4  
K: C  
Q: 1/4=60  
L: 1/8  
dc Ad | dc Ad | dc Ad | GG AA |  
w:Ra-ul ti-nhaum i-oi-o que io-io-ia-va to-do o dia  
z/2 D/2E/2F/2 G>G | AG F2 | 1 z/2 D/2E/2F/2 G>G | AG F2 :|2 z F/2A/2 GG | FE DD |]  
w:quan-doo Ra-úl fa-zia ó-ó o i-o-io a-dor-me-cia  
</abc>

Raúl tinha um ioio  
que ioioiava todo o dia  
quando o Raúl fazia ó-ó  
o ioio adormecia

E quando o Raúl chorava  
porque o ó-ó não vinha  
o ioio embalava  
para baixo e para cima

Raúl dormia e sonhava  
e quando sonhava sorria  
porque o io-io ioioiava  
nos sonhos que Raúl via

### Output teste 3

```
\title{ = Raúl tinha um Ioio}
\author{ Manuel António Pina, Suzana Ralha}
\documentclass[12pt]{article}
\begin{document}
\maketitle
\section*{Letra}
\begin{center}
\begin{verbatim}
Raúl tinha um ioio
que ioioiava todo o dia
quando o Raúl fazia ó-ó
o ioio adormecia

E quando o Raúl chorava
porque o ó-ó não vinha
\end{verbatim}


```

o ioio embalava  
para baixo e para cima

Raúl dormia e sonhava  
e quando sonhava sorria  
porque o io-io ioioiava  
nos sonhos que Raúl via  
\end{verbatim}  
\vspace{5mm}  
\hfill Bando dos Gambozinos  
\end{center}  
\end{document}

# Amndoaa Amarga

Jos Carlos Ary dos Santos, Alain Oulman

April 1, 2015

## Letra

Port ti falo  
e ningum pensa  
mas eu digo  
minha amndoaa, meu amigo  
meu irmo  
meu tropel de ternura  
minha casa  
meu jardim de carncia  
minha asa.

Por ti vivo  
e ningum pensa  
mas eu sigo  
um caminho de silvas  
e de nardos  
uma intensa ternura  
que persigo  
rodeada de cardos  
por tantos lados.

Por ti morro  
e ningum sabe  
mas eu espero  
o teu corpo que sabe  
a madrugada

Figura 6.13: PDF gerado por o ficheiro latex (teste 1). Pagina 1 de 2

Por ti morro  
e ningum sabe  
mas eu espero  
o teu corpo que sabe  
a madrugada

1

---

o teu corpo que sabe  
a desespero

minha amarga amnhoa  
desejada.

minha amarga amnhoa  
desejada.

Amlia Rodrigues

Figura 6.14: PDF gerado por o ficheiro latex (teste 1). Pagina 2 de 2

# \* Tejo que levas as guas

Manuel da Fonseca, Adriano Correia de Oliveira

April 1, 2015

## Letra

Tejo que levas as guas  
correndo de par em par  
lava a cidade de mgoas  
leva as mgoas para o mar

Lava-a de crimes espantos  
de roubos, fomes, terrores,  
lava a cidade de quantos  
do dia fingem amores

Leva nas guas as grades  
de aço e silêncio forjadas  
deixa soltar-se a verdade  
das bocas amordaoadas

Lava bancos e empresas  
dos comedores de dinheiro  
que dos salários de tristeza  
arrecadam lucro inteiro

Lava palácios vivendas  
casebres bairros da lata  
leva negócios e rendas  
que a uns farta e a outros mata

Tejo que levas as guas  
correndo de par em par  
lava a cidade de mgoas  
leva as mgoas para o mar

Lava avenidas de vrios  
vielas de amores venais  
lava albergues e hospcios  
cadeias e hospitais

Afoga empenhos favores  
vs glrias, ocas palmas  
leva o poder dos senhores  
que compram corpos e almas

Leva nas guas as grades  
...

Das camas de amor comprado  
desata abraos de lodo  
rostos corpos destroados  
lava-os com sal e iodo

Tejo que levas nas guas  
...

Adriano Correia de Oliveira

# = Ral tinha um Ioio

Manuel Antnio Pina, Suzana Ralha

April 1, 2015

## Letra

Ral tinha um ioio  
que ioioiava todo o dia  
quando o Ral fazia -  
o ioio adormecia

E quando o Ral chorava  
porque o - no vinha  
o ioio embalava  
para baixo e para cima

Ral dormia e sonhava  
e quando sonhava sorria  
porque o io-io ioioiava  
nos sonhos que Ral via

Bando dos Gambozinos

Figura 6.17: PDF gerado por o ficheiro latex (teste 3)