

# Trabalho Prático 1

Processamento de Linguagens  
Universidade do Minho

André Peixoto (a82260)  
Filipe Cunha (a83099)  
Luis Martins(a82298)

Março 2019



# Índice

<b>1</b>	<b>Introdução</b>	<b>3</b>
<b>2</b>	<b>Autómatos</b>	<b>4</b>
2.1	Autómato - ficheiro TXT . . . . .	4
2.1.1	Expressões Regulares . . . . .	4
2.1.2	Representação . . . . .	5
2.2	Autómato - ficheiro SQL . . . . .	5
2.2.1	Expressões Regulares . . . . .	5
2.2.2	Representação . . . . .	5
<b>3</b>	<b>Demonstração</b>	<b>7</b>
3.1	Funcionalidades . . . . .	7
3.2	Resultado Final . . . . .	7
<b>4</b>	<b>Conclusão</b>	<b>14</b>

# 1 Introdução

Neste relatório, descrevemos em detalhe a nossa resolução do primeiro trabalho prático de Processamento de Linguagens. Este tem como objetivo aprofundar os nossos conhecimentos sobre Flex e ganhar aptidão a processar ficheiros de texto.

Para isto, dos 8 enunciados apresentados, e conforme o cálculo da resto da divisão inteira, o enunciado que nos ficou designado foi o 4º, "Wikipedia PT, language links".

Os requisitos deste baseiam-se, principalmente, na especificação de padrões que vêm dentro dos ficheiros input através de expressões regulares e, adicionalmente, pela construção de um filtro que consegue os reconhece e processa de acordo com o que pretendemos.

## 2 Autómatos

O grupo, para realizar o trabalho, teve que se basear em autómatos, que serviram como base de pensamento de como iríamos executar a leitura dos ficheiros através do *flex*. Estes são de relevante importância, visto que ajudam a compreender o mecanismo em que esta se baseou. À medida que se avançou no trabalho, também se teve que corrigir os autómatos, uma vez que nos fomos deparando com exceções que não queríamos que fossem guardadas nos nossos termos em português. Por outro lado, os termos portugueses, que são guardados, são de extrema importância para sabermos os termos que serão guardados das outras línguas.

Apresentamos, nesta secção, os dois autómatos em que nos baseamos. Um para interpretar o ficheiro txt que contém os termos portugueses e o outro para interpretar o ficheiro sql, sendo que sem a leitura do primeiro não seria possível executar o segundo.

### 2.1 Autómato - ficheiro TXT

#### 2.1.1 Expressões Regulares

Condição	Expressão Regular
INT	[< title >]
ENDT	[< "/"title >]
D	[0-9 +]
FAIL 1	(.*\n*)[( D+ [-]D+ )]
FAIL 2	.*[( D+ )]
FAIL 3	((.*E)+ (E.*)+)
FAIL 4	W.*
IGNORE 1	(Cat.*)
Cat	Categoria:
IGNORE 2	(Port.*)
Port	Portal:
IGNORE 3	(Prede.*)
Prede	Predefinição:
IGNORE 4	(Ajuda.*)
Ajuda	Ajuda:

### 2.1.2 Representação

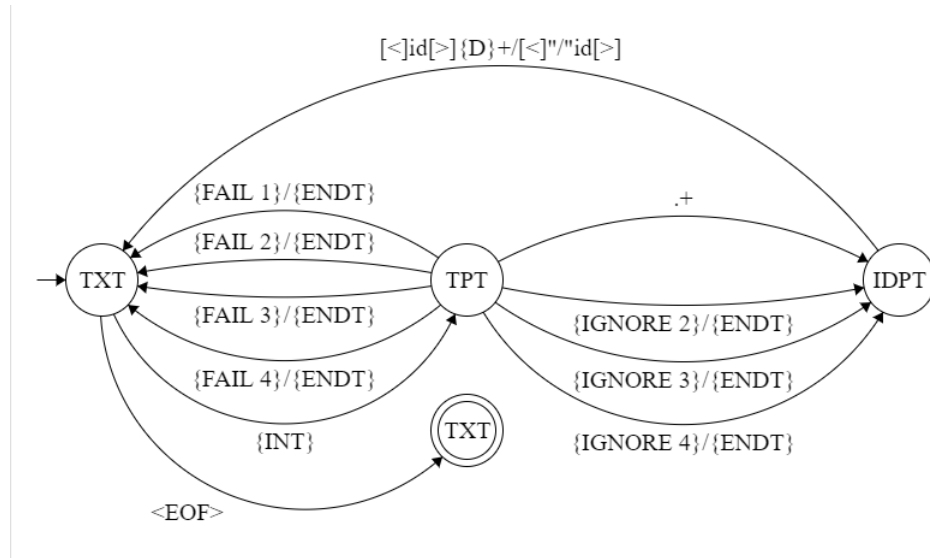


Figure 1: Autómatos TXT

## 2.2 Autómatos - ficheiro SQL

### 2.2.1 Expressões Regulares

Condição	Expressão Regular
x	de/en/es/it/fr/ru

### 2.2.2 Representação

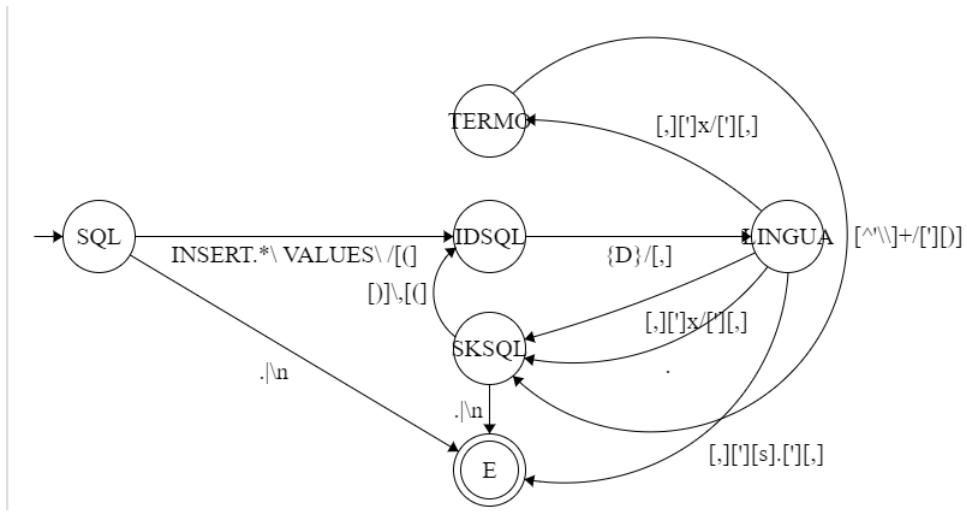


Figure 2: Autómato SQL

### 3 Demonstração

Nesta secção, demonstramos as funcionalidades que implementamos na resolução do nosso enunciado, a partir dos autómatos em que nos baseamos, assim como todo o resultado final correspondente a uma leitura/processamento dos ficheiros txt e sql.

#### 3.1 Funcionalidades

Para além dos requisitos base à solução do problema, desenvolvemos também umas funcionalidades extra. Estas, assim como a utilização do processador de texto, estão explícitas na seguinte figura:

```
-----HELP-----
Utilização:
make
./tp1 [Nº de tuplos a serem processados-opcional] [nome].txt [nome].sql
-----
OUTRAS FUNCIONALIDADES:
0 -> Terminar o programa
1 -> Dado um id de um termo, dar a tradução numa dada linguagem
2 -> Dado um id de um termo, apresentar todas as traduções
3 -> Vizualizar tabela de tradução produzida
4 -> Perguntar qual a linguagem com mais tradução
-----HELP-----
```

Figure 3: Funcionalidades

#### 3.2 Resultado Final

```
Digite a opção pretendente: 1
Id do termo a ser traduzido: 1054

Linguagem pretendida: FR

Termo (PT): Indústria química
Termo (FR): Industrie chimique
```

Figure 4: Funcionalidade 1

```

Digite a opção pretendente: 2
Id do termo a ser traduzido: 1054
Resultado:

Termo (PT): Indústria química
Termo (DE): Chemische Industrie
Termo (EN): Chemical industry
Termo (ES): Industria química
Termo (FR): Industrie chimique
Termo (IT): Industria chimica
Termo (RU): Химическая промышленность

```

Figure 5: Funcionalidade 2

```

Digite a opção pretendente: 3
IdTermo , PT , DE , EN , ES , FR , IT , RU
220, Astronomia,Astronomie,Astronomy,Astronon
223, América Latina,Lateinamerika,Latin Ameri
224, Albino Forjaz de Sampaio,,,,,,,,
226, Anno Domini,Anno Domini,Anno Domini,Ann
228, Aquiles,Achilleus,Achilles,Aquiles,Achil
229, Anarcocapitalismo,Anarchokapitalismus,Ar
нархо-капитализм,
230, Anarquismo,Anarchismus,Anarchism,Anarqu
231, Albert Einstein,Albert Einstein,Albert E
232, Aquecimento global,Globale Erwärmung,Glo
bale,Глобальное потепление,
234, Antraz,,,,,,,,
235, Lista de padrões de arquivo gráfico,Grat
236, Abel (desambiguação),Abel,Abel (disambig
237, Adams,Adams,Adams,Adams,Adams,Adams,Ad
238, Adriano,Hadrian (Kaiser),Hadrian,Adriano
239, Alberto,Alberto,Alberto,Alberto (desambi
240, Alexandre (nome),,Alexandre (given name)
241, Alexis,,,,,,,,
242, Allen,Allen,Allen,Allen (desambiguación)
243, Afonso de Aragão,,Alfonso of Aragon,Alfo
244, Afonso, Príncipe de Portugal (1475-1491)
(1475-1491),Alphonse de Portugal (1475-1491)

```

Figure 6: Funcionalidade 3



```
Digite a opção pretendente: 4
A língua que mais termos PT traduziu é a DE
```

Figure 7: Funcionalidade 4

**Lingua: PT**



**Estatística:**

Número de Termos selecionados para tradução: 2295641  
Número de Termos analisados (termos totais): 2405447  
Media de traduções por palavra: 1.654  
Número de entradas ignoradas: 109806  
Número de prefixos removidos: 421592

**Entradas Removidas:**

Anos na forma (\*\*\*\*): 7315  
Anos na forma (\*\*\*\*-\*\*\*\*): 1308  
Anos na forma \*\*\*\* a.C.: 4505  
Wikipedia: 94502  
! : 2176

**Prefixos Removidos:**

Categoria: 321680  
Portal: 21727  
Predefinição: 77853  
Ajuda: 332

[voltar](#)

Figure 8: Ficheiro HTML PT

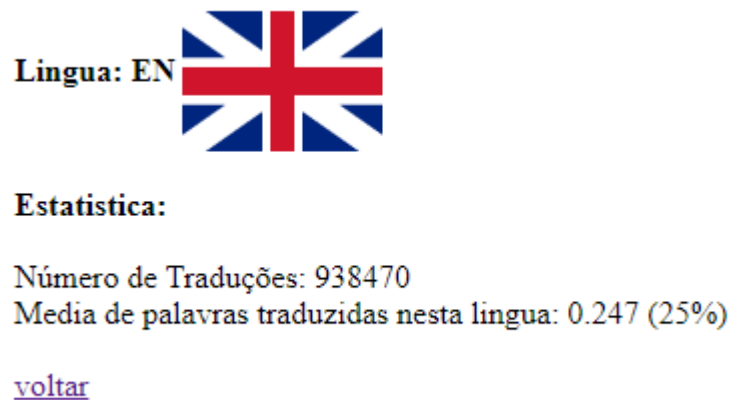


Figure 9: Ficheiro HTML EN

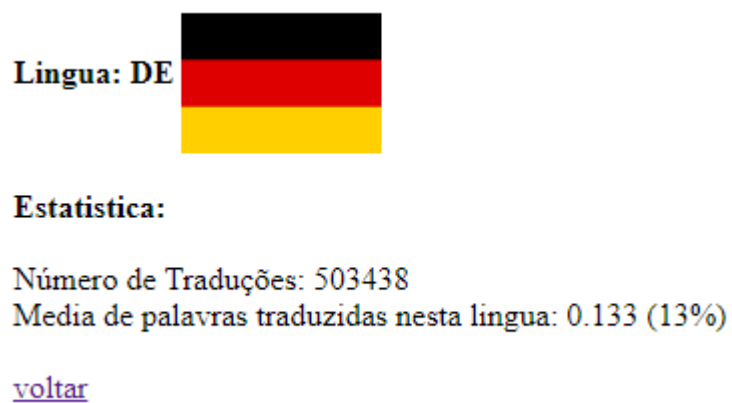


Figure 10: Ficheiro HTML DE

**Lingua: ES**



**Estatística:**

Número de Traduções: 617486

Media de palavras traduzidas nesta lingua: 0.163 (16%)

[voltar](#)

Figure 11: Ficheiro HTML ES

**Lingua: FR**



**Estatística:**

Número de Traduções: 576593

Media de palavras traduzidas nesta lingua: 0.152 (15%)

[voltar](#)

Figure 12: Ficheiro HTML FR

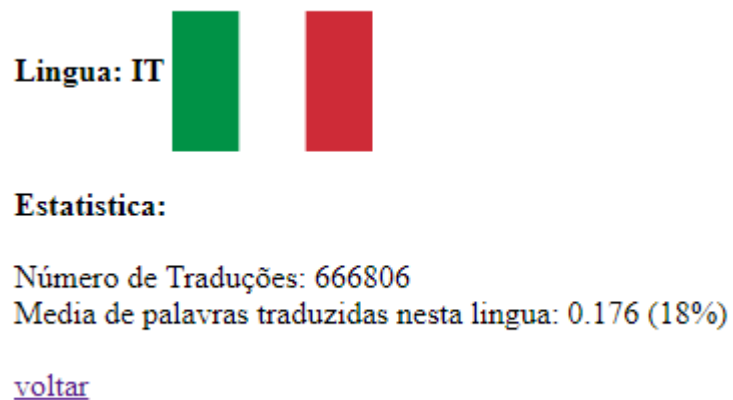


Figure 13: Ficheiro HTML IT

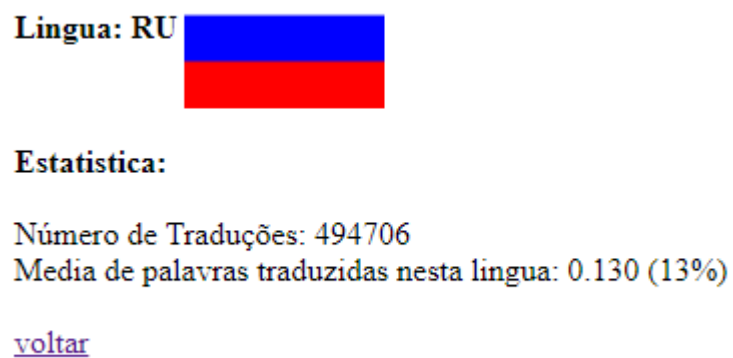


Figure 14: Ficheiro HTML RU

idTermo	PT	DE	EN	ES	FR	IT	RU
220	Astronomia	Astronomie	Astronomy	Astronom	Astronomie	Astronomia	Астрономия
223	América Latina	Lateinamerika	Latin America	América	Amérique	America Lat	Латинская Америка
224	Albino Forjaz de Sampaio						
226	Anno Domini	Anno Domini	Anno Domini	Anno Dom	Anno Domini	Anno Domini	Анно Домини
228	Aquiles	Achilleus	Achilles	Aquiles	Achille	Achille	Ахиллес
229	Anarcocapitalis	Anarchokapital	Anarcho-capit	Anarcocap	Anarcho-cap	Anarco-capit	Анархокапитализм
230	Anarquismo	Anarchismus	Anarchism	Anarquism	Anarchisme	Anarchismo	Анархизм
231	Albert Einstein	Albert Einstein	Albert Einstei	Albert Ein	Albert Einste	Albert Einste	Альберт Эйнштейн
232	Aquecimento global	Globale Erwärmung	Global warmi	Calentami	Réchauffe	Riscaldamento	Глобальное потепление
234	Antraz						Антракс
235	Lista de padrões	Grafikformat	Image file formats				Список форматов изображений
236	Abel (desambiguação)	Abel	Abel (disambiguation)	Abel (desambiguação)	Abel (homonym)	Abel	Абель (дисквалификация)
237	Adams	Adams	Adams	Adams	Adams	Adams	Адамс
238	Adriano	Hadrian (Kaiser)	Hadrian	Adriano	Hadrien	Adriano	Адриано
239	Alberto	Alberto	Alberto	Alberto (disambiguation)	Alberto	Alberto (disambiguation)	Альберто
240	Alexandre (nome)		Alexandre (given name)	Alexandre (disambiguation)	Alexandre (prénom)		Александр (имя)
241	Alexis						Алексис
242	Allen	Allen	Allen	Allen (disambiguation)	Allen	Allen	Аллен
243	Afonso de Aragão		Alfonso of Aragon	Alfonso de Aragon		Aragona	Афонсо де Арагон
244	Afonso	Príncipe de Fátima	Infant vor Afonso			Prince of Portugal	Афонсо де Португалия

Figure 15: Tabela CSV

## 4 Conclusão

Ao longo deste projeto, conseguimos ultrapassar várias etapas e, para tal, foi necessário estudar e aprofundar os nossos conhecimentos à cerca da linguagem de filtragem de texto Flex, a fim de desenvolvermos bem a solução do nosso enunciado.

Para concluir, podemos dizer que estamos satisfeitos com o nosso trabalho, pois conseguimos responder corretamente ao que nos foi proposto. Assim, é da nossa opinião que este foi um trabalho prático muito relevante e útil para o nosso percurso académico, na medida em que nos levou a adquirir bastante experiência nesta área de processamento de linguagens.