



Universidade do Minho
Escola de Engenharia

INTRODUÇÃO A PROCESSAMENTO
DE
LINGUAGENS NATURAIS

**Portuguese Grammar
Processor(PGP)**

João Costeira, A78073
Bernardo Silva, A77230
Paulo Jorge, A78203

Contents

1	Introdução	2
2	Dependências da Aplicação	2
3	Descrição da Solução	2
3.1	Descrição da aplicação	2
4	Funcionalidades e Exemplos de Execução	3
4.1	Invocação sem <i>flags</i>	3
4.2	Opção <i>-w</i>	4
4.3	Opção <i>-r</i>	5
4.4	Opção <i>-n</i>	6
4.5	Opção <i>-v</i>	7
4.6	Opção <i>-p</i>	7
4.7	Mútua exclusividade	9
5	Conclusão	9
6	Referências	9

1 Introdução

Para o terceiro trabalho prático, o grupo decidiu o desenvolvimento de um programa inspirado no tema 1 sugerido pelo professor, o desenvolvimento de um processador de gramática portuguesa, onde sobre um conjunto de palavras consegue identificar a sua morfologia e relações existentes.

Neste relatório encontra-se descrito as dependências da aplicação e o seu funcionamento geral.

2 Dependências da Aplicação

De forma a executar o programa é necessário a instalação das seguintes bibliotecas *python*:

Dependências da Aplicação	
<i>nltk</i>	Ficheiro fornecido pelos professores que permite a identificação da morfologia do texto
<i>pickle</i>	Carregamento de ficheiros
<i>arg Parser (pip install argparse)</i>	Usado no tratamento de opções (<i>flags</i>) da aplicação
<i>pyvis (pip install pyvis)</i>	Visualizador de relações (<i>network</i>)

Table 1: Descrição do problema

3 Descrição da Solução

Nesta secção encontra-se descrita a estrutura do programa e em que sentido essas diferentes estruturas são utilizadas de forma a implementar as funcionalidades da aplicação.

3.1 Descrição da aplicação

A seguinte tabela contém para cada um dos ficheiros que constituem a aplicação, uma breve descrição da sua funcionalidade.

Descrição da Aplicação	
<i>args.py</i>	Tratamento das opções (<i>flags</i>) da aplicação e tratamento de opções mutuamente exclusivas
<i>treino.py</i>	Responsável por gerar o ficheiro <i>mac_morpho.plk</i> , utilizado na identificação morfológica das palavras
<i>graphics.py</i>	Responsável por gerar gráficos em ficheiros <i>.html</i> com relações de palavras existentes no programa
<i>pgp</i> ou <i>comando.py</i>	Programa principal que efetua o processamento do texto de acordo com as opções passadas

Table 2: Modelação dos utilizadores

4 Funcionalidades e Exemplos de Execução

Com a invocação do comando com as *flags* de *help*:

```

1 $pgp -h
2 $pgp --help

```

Podemos observar todas as funcionalidades que o programa possui.

4.1 Invocação sem *flags*

A invocação do programa sem *flags* permite a identificação morfológica de cada uma das palavras passadas pelo *stdin* ao programa.

Por exemplo:

```
1 $pgp
2 ola
3 tudo
4 bem
5 $ctrl-d
```

Obtém o seguinte resultado:

```
1 ('ola', 'N')
2 ('tudo', 'PROSUB')
3 ('bem', 'IN')
```

Com recurso à *flag* *-i* ou *-input* é possível a identificação morfológica sobre um ficheiro passado como argumento.

```
1 $pgp -i os_maias.txt
2 $pgp --input os_maias.txt
```

Obtemos resultados semelhantes aos anteriores, mas neste caso sobre a obra completa *Os Maias* de Eça de Queiroz.

4.2 Opção *-w*

A opção *-w* ou *word* permite efetuar pesquisas no texto de palavras de acordo com a identificação dada pelo *nltk*.

Por exemplo podemos efectuar a pesquisa de todos os nomes existentes na obra da seguinte forma:

```
1 $pgp -i os_maias.txt -w N
2 $pgp -i os_maias.txt -w NPROP -o out.txt
```

4.3 Opção *-r*

A opção *-r* ou *regEx* permite a pesquisa de triplos no ficheiro de acordo com expressões regulares.

Como esta opção envolve relações de palavras existentes numa obra, é possível a geração de grafos das relações encontradas.

De forma exemplificativa:

```
1 $pgp -i os_maias.txt -r Carlos \\w+ \\w+
2 $pgp -i os_maias.txt -r Carlos [a-z]+ [a-z]+
3 $pgp -i os_maias.txt -r Carlos \\w+ \\w+ -g grafo.html
```

Após a execução do comando com a opção *-g* (gerar grafos), podemos visualizar as relações de palavras inicializadas por Carlos (triplos Carlos palavra palavra):

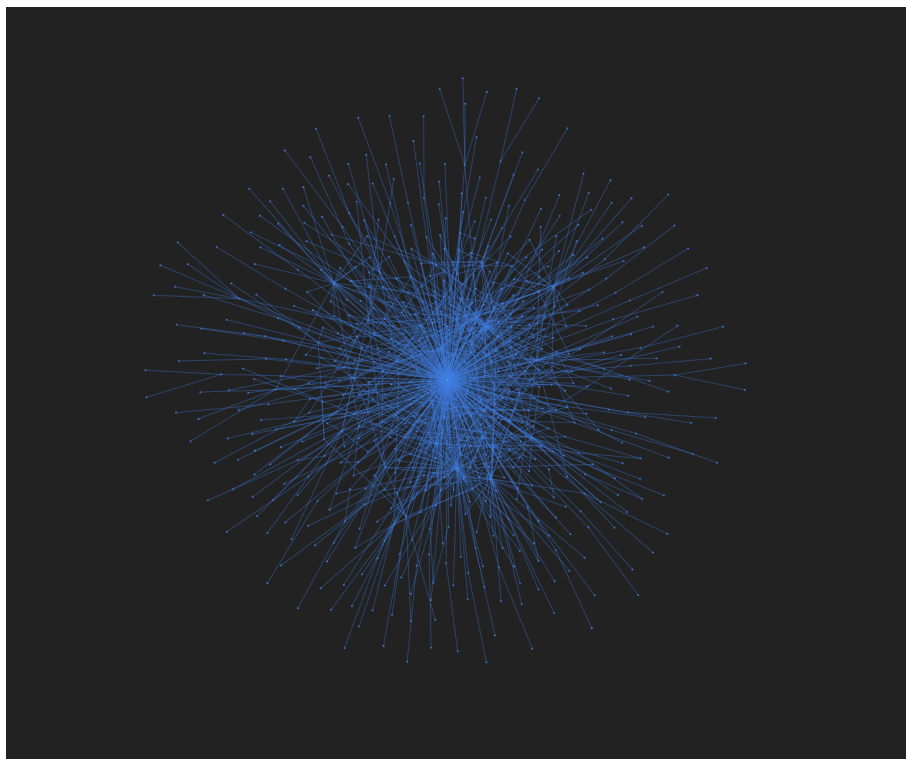


Figure 1: Grafo com as Relações centradas em Carlos

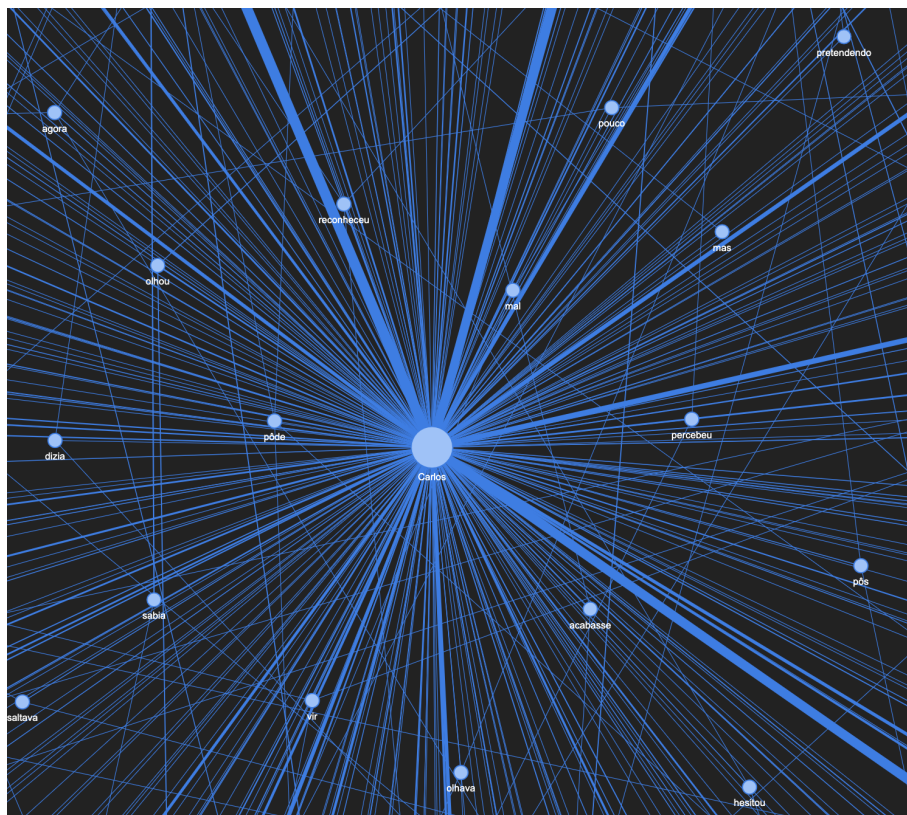


Figure 2: Podemos visualizar que as relações do grafo encontram-se centradas na personagem Carlos

4.4 Opção *-n*

A opção *-n* ou nomes permite encontrar as relações entre nomes existentes na mesma frase.

Semelhante ao caso anterior, como estamos perante relações, é possível a geração de grafos para visualizar as relações entre as palavras.

```
1 $pgp -i os_maias.txt -n
2 $pgp -i os_maias.txt -n -g nomes.html
```

4.5 Opção *-v*

Com a opção *-v* ou verbos, podemos visualizar as relações de palavras centradas por um verbo, assim triplos de palavras são gerados.

Exemplos de execução:

```
1 $pgp -i os_maias.txt -v
2 $pgp -i os_maias.txt -v -g ver.html
3 $pgp -i os_maias.txt -v -o ver.txt
```

4.6 Opção *-p*

A opção *-p* ou de palavra permite efectuar procuras triplos no texto que segue o seguinte padrão: (Agente_da_Relação,verbo,palavra). Desta forma procuras sobre frases de uma obra que seguem o padrão "*Maria ... era.. palavra*" ou "*Maria... gosta ... palavra*" pode ocorrer.

```
1 $pgp -i os_maias.txt -p Maria
2 $pgp -i os_maias.txt -p Maria -o maria.txt
3 $pgp -i os_maias.txt -p Maria -g maria.html
```

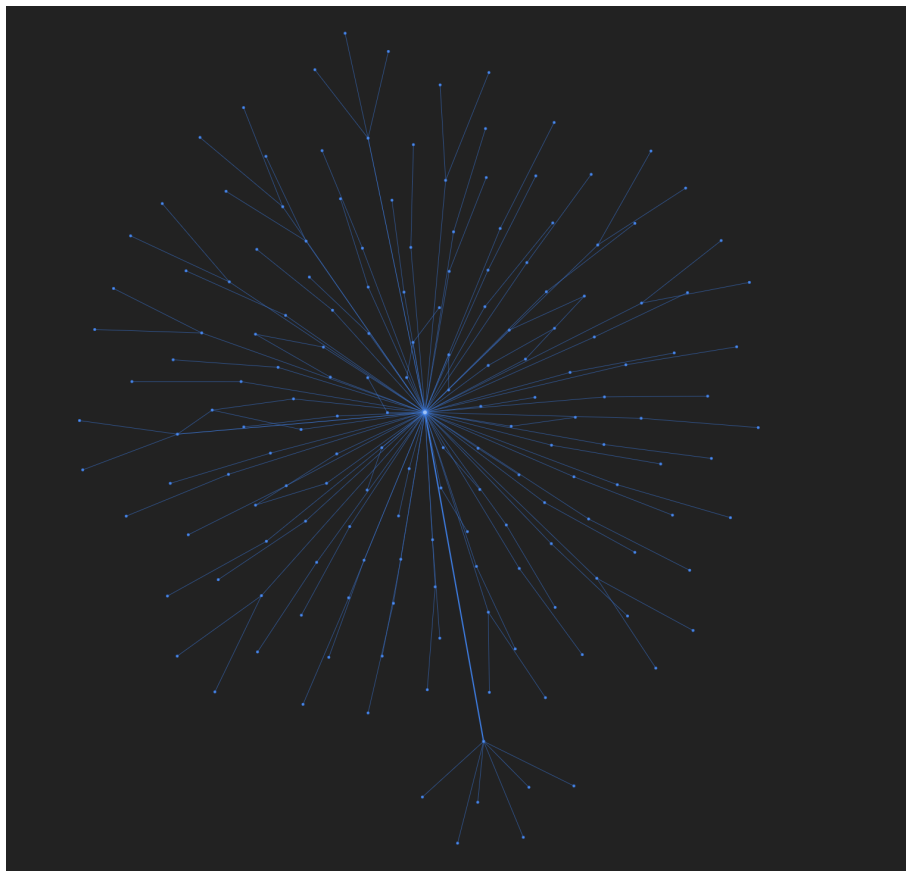


Figure 3: Grafo com as relações das frase centradas em Maria

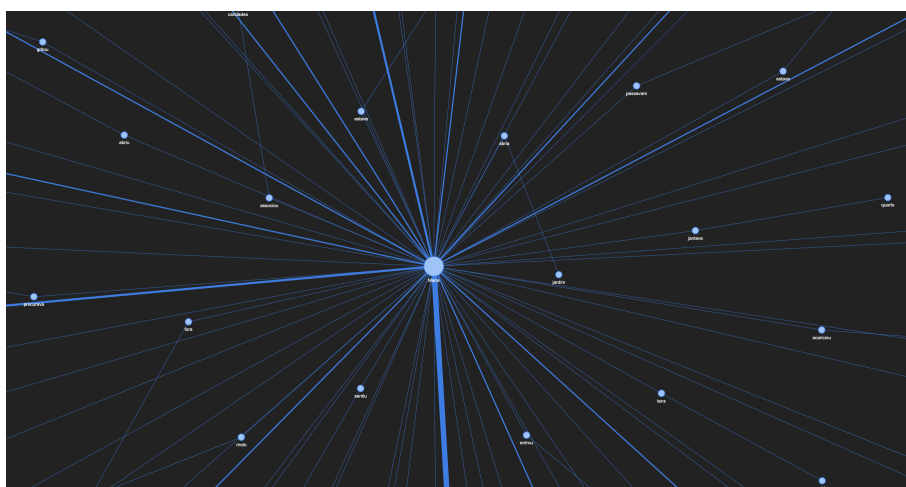


Figure 4: Mesmo grafo com mais detalhe

4.7 Mútua exclusividade

De realçar que sobre as opções foi tomado em consideração os seguintes cenários:

- Invocações que geram múltiplos resultados - Caso seja efectuada uma invocação que por exemplo pede o *output* em ficheiro (-o) e gerar o grafo (-g), este caso provoca dois resultados. Como o objectivo deste programa é usá-lo como um comando, execuções que geram múltiplos *outputs* são proibidas, informando o utilizador da sua múltipla exclusividade.
- Grafos - A geração de grafos só é permitida em invocações que geram relações entre palavras. Desta forma invocações que simplesmente procuram todos os nomes ou verbos do ficheiro e simultaneamente pedem a geração de grafos sobre estas palavras não relacionadas são proibidas. O utilizador é indicado sobre a sua exclusividade mutua.
- Argumentos - Opções como procura de triplos no texto de acordo com expressões regulares, para além da *flag* que indica seu uso, necessita de três argumentos extra (cada uma das expressões regulares). Caso estrs não sejam adicionados, a execução não ocorre e o utilizador é avisado.

5 Conclusão

Com este trabalho foi possível desenvolvimento de uma aplicação que permite efectuar o processamento de relações de palavras num texto.

De acordo com as *flags* indicadas, o grupo acredita que desenvolveu uma aplicação de acordo com o enunciado e funcionalidades extra.

6 Referências

- <https://pyvis.readthedocs.io/en/latest/tutorial.html>
- <https://docs.python.org/3/library/argparse.html>
- <https://github.com/andrefs/ipln-1920-i>