

Quotation Extractor for Portuguese

Manual do Usuário

Rafael Reis
`rrsilva@inf.puc-rio.br`
Orientador: Ruy Milidiú

9 de Dezembro de 2015

Conteúdo

1	Apresentação	2
2	Requisitos Técnicos	2
3	Instalação	2
4	Configuração	2
5	Execução	2
6	Testes	3

1 Apresentação

Este manual contém as instruções para a utilização do Quotation Extractor.

2 Requisitos Técnicos

Para utilizar o programa, é necessário ter o Python 3.5.0 instalado. Recomenda-se a distribuição Anaconda¹, que já possui os módulos `numpy` e `scikit-learn`.

É altamente recomendável a utilização de um sistema UNIX (Linux, Mac OS etc.) para o acesso por *bash*/terminal.

3 Instalação

Copie a pasta `qextractor` do CD para uma outra pasta num disco que permita escrita. Chamaremos essa outra pasta de **HOME**.

4 Configuração

O programa já possui as bases de treino e de teste GLOBOQUOTES instaladas. Porém, você pode alterá-las, desde que elas estejam no formato GLOBOQUOTES, a seguir:

```
[word, pos, np, ne, ck, clause, gpq, coref, gquotestart, gquoteend, gquote]
```

Para utilizar as suas bases, substitua os arquivos em:

- 1 HOME/qextractor/data/corpus-globocom-cv.txt
- 2 HOME/qextractor/data/corpus-globocom-test.txt

onde 1 é o caminho para a base de treino e 2 o caminho para a base de teste.

5 Execução

Para executar o programa, basta rodar o comando:

```
python -m qextractor.application
```

Ele exibirá como saída o modelo *w* calibrado no arquivo de treino e o resultado da aplicação do modelo no arquivo de teste.

¹<https://www.continuum.io/downloads>

6 Testes

Para a execução de todos os testes unitários utilizando o *framework* `unittest` do Python, basta copiar o arquivo `Makefile` para a pasta **HOME** e utilizar a opção `test` do *make*:

```
make test
```