

# Collecting Data About Universities

Programming Languages

6170 – Bruno Moreira



Instituto Politécnico de Beja

Escola Superior de Tecnologia e Gestão

- O trabalho tem como objectivo a obtenção e tratamento de dados existentes tanto em páginas html como em base de dados. Neste caso foram utilizados os dados existentes na página rebides, <http://www.rebides.oces.mctes.pt/rebides>\*, que contem dados relativos ao ensino superior do ano 2000 ao ano 2009.
- Os dados retirados da página acima mencionada, foram posteriormente carregados em base de dados.
- O tratamento dos dados, consiste em efectuar listagens e estatísticas de uma forma dinâmica recorrendo à linguagem de programação Python.

## Resumo



# Programa

Descrição das funcionalidades e classes



- O trabalho tem como objectivo a obtenção e tratamento de dados existentes tanto em páginas html como em base de dados. Neste caso foram utilizados os dados existentes na página rebides, <http://www.rebides.oces.mctes.pt/rebides>\*, que contem dados relativos ao ensino superior do ano 2000 ao ano 2009.
- Os dados retirados da página acima mencionada, foram posteriormente carregados em base de dados.
- O tratamento dos dados, consiste em efectuar listagens e estatísticas de uma forma dinâmica recorrendo à linguagem de programação Python.

## Programa



### ▪ **OBTENÇÃO DE DADOS**

Esta funcionalidade consiste na obtenção dos dados da página online, trata-los por forma a retirar apenas o que interessa, ou seja os dados referentes aos estabelecimentos e docentes.

### ▪ **EXTRACÇÃO DE LISTAS**

Foi-nos solicitado que, com base nos dados acima mencionados, fossem disponibilizadas algumas listagens. Efectuei alguns melhoramentos, por forma a ser possível efectuar a extracção de listagens de uma forma dinâmica, ajustando-se melhor às necessidades dos utilizadores.

## Funcionalidades



#### ▪ **EXTRACÇÃO DE ESTATÍSTICAS**

Foi-nos solicitado que, com base nos dados acima mencionados, fossem disponibilizadas algumas estatísticas. Efectuei alguns melhoramentos, por forma a ser possível efectuar a extracção de estatísticas de uma forma dinâmica, ajustando-se melhor às necessidades dos utilizadores.

#### ▪ **EXTRACÇÃO DE FICHEIROS CSV**

As consultas e estatísticas acima mencionadas permitem efectuar a extracção dos dados para ficheiros no formato CSV, facilitando assim a importação para outras plataformas, por exemplo Excel.

#### ▪ **DISPONIBILIZAÇÃO EM HTML**

Toda a informação extraída, é disponibilizada em formato HTML, através de um servidor HTTP simples.

## Funcionalidades



## Classes

- **BASEDADOS.PY**

Esta classe é responsável pelas interacções activas na base de dados, ou seja todas as interacções que podem de alguma forma alterar a estrutura e ou conteúdo da mesma. Para tal foi utilizada a biblioteca sqlite3.

- **CRIARCSV.PY**

Esta classe é responsável pela criação de ficheiros CSV (Comma-separated values), com base em dados resultantes das consultas e estatísticas solicitadas pelo utilizador. Foi utilizada a biblioteca csv.

- **DADOS.PY**

Esta classe é responsável pelos acessos efectuados à base de dados, numa óptica de efectuar pesquisas ou consultas à mesma, utilizando para tal as bibliotecas sqlite3.

- **GRAPHS.PY**

Esta classe é responsável pela disponibilização de gráficos com base nos dados existentes em base de dados, utilizando para tal a biblioteca matplotlib.

- **HTML.PY**

Esta classe é responsável pela transformação dos dados existentes em base de dados em páginas HTML. Foram utilizadas as bibliotecas shutil e os, para criar, eliminar e validar existência de ficheiros e ou directórios.

- **HTTPSERVER.PY**

Esta classe é responsável pela disponibilização dos conteúdos em formato HTML utilizando as bibliotecas SimpleHttpServer e SocketServer. Apoia-se na classe KThread.py para permitir que o servidor HTTP corra em paralelo com a aplicação.

## Classes





- **KTHREAD.PY**

Esta classe é responsável pela criação de processamento em processos separados, permitindo assim soltar o processo de origem para efectuar outras operações.

- **MAIN.PY**

Esta é a classe responsável por fazer a ponte entre a camada applicacional do programa e a sua interface, permitindo assim alterar a interface sem ter que alterar a camada applicacional.

- **OBTERDADOS.PY**

Esta classe é responsável pela obtenção dos dados em formato HTML que existem na página Rebides, utilizando a biblioteca BeautifulSoup que permite fazer o parse das várias tag's HTML e tratar a informação que realmente interessa de uma forma eficaz.

- **REBIDES.PY**

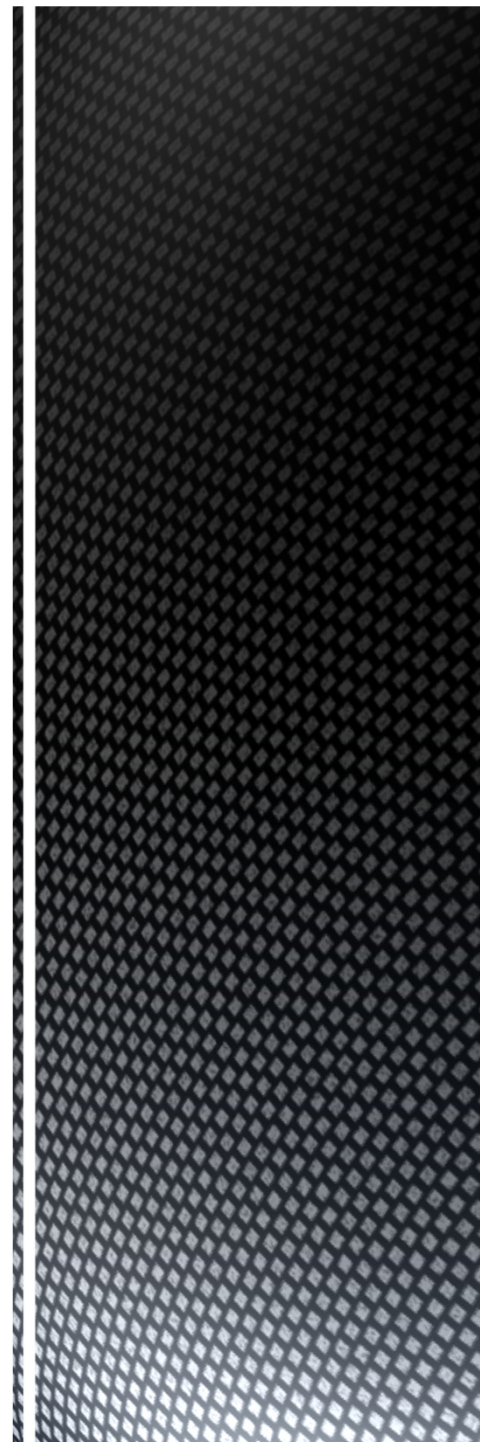
Esta classe é responsável pela disponibilização da interface do programa. Esta foi construída utilizando a aplicação wxglade e utiliza a biblioteca wx.

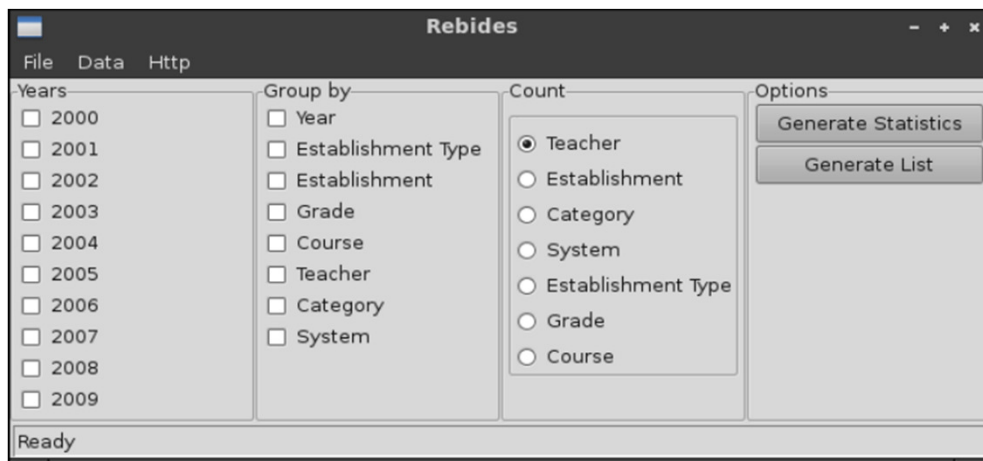
## Classes



# Interface

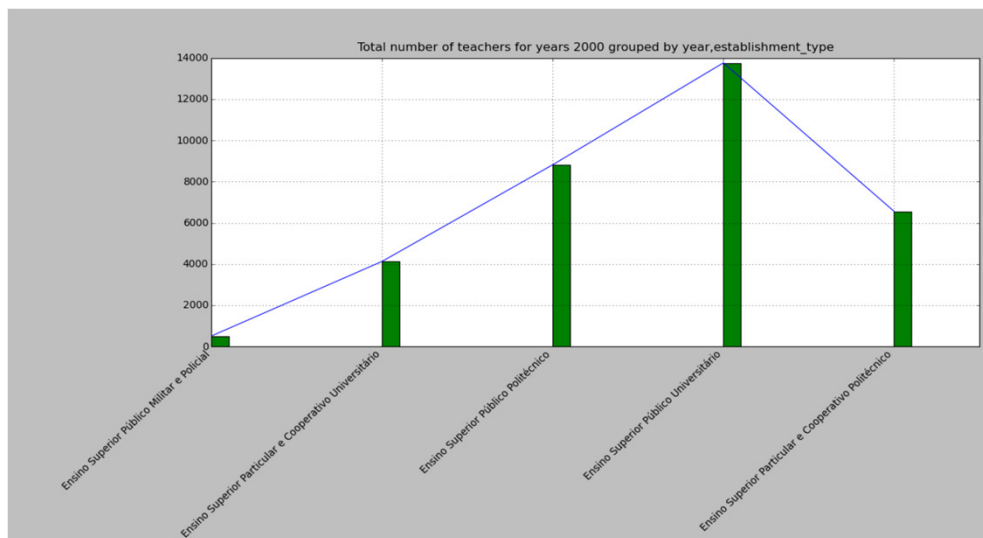
Screenshots da interface a listagens/gráficos





## Interface

Onde podemos operar todo o programa



## Gráfico

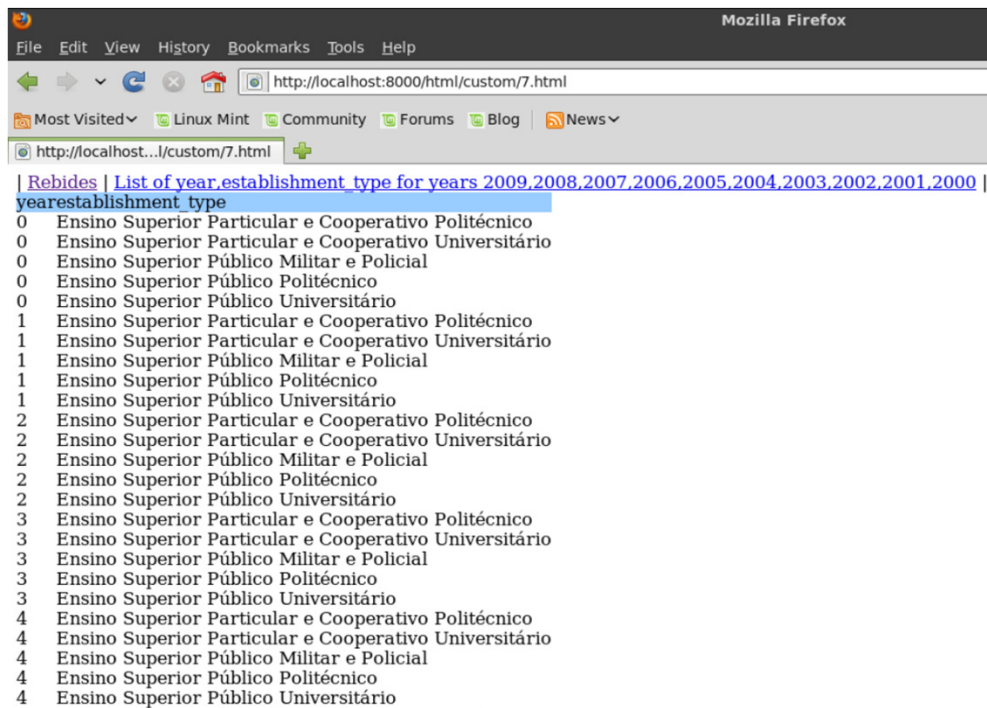
Exemplo de gráfico extraído da aplicação

```
|-----  
"Total number of teachers for years 2000 grouped by year,establishment_type"  
-----  
0,Ensino Superior Público Militar e Policial,509  
0,Ensino Superior Particular e Cooperativo Universitário,4129  
0,Ensino Superior Público Politécnico,8821  
0,Ensino Superior Público Universitário,13765  
0,Ensino Superior Particular e Cooperativo Politécnico,6558
```

## CSV

Exemplo de CSV extraído da  
aplicação





## HTML

Exemplo de página HTML  
extraída da aplicação

Fim da apresentação

