

Instituto Politécnico de Beja

Escola Superior de Tecnologia e Gestão



# Collecting Data About Universities

Programming Languages

Bruno Moreira

11

Docente: José Jasnau Caeiro

Aluno: Bruno Alexandre da Silva Moreira

Número: 6170

Curso: Eng<sup>a</sup> Informática 2º ano



## ÍNDICE

Resumo.....	3
Programa.....	4
Definição geral de funcionalidades .....	4
Obtenção de dados .....	4
Extracção de listas .....	4
Extracção de Estatísticas .....	4
Extracção de ficheiros CSV .....	4
Disponibilização em HTML .....	4
Descrição de classes .....	5
BaseDados.py .....	5
CriarCSV.py .....	5
Dados.py .....	5
Graphs.py .....	5
Html.py .....	5
HttpServer.py .....	5
KThread.py .....	5
Main.py.....	5
ObterDados.py .....	6
Rebides.py .....	6
Conclusão .....	7
Bibliografia .....	8
Anexos .....	9
Interface .....	9
HTML .....	9
CSV .....	9

## RESUMO

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

### DEFINIÇÃO GERAL DE FUNCIONALIDADES

Neste trabalho foi-nos solicitado que o programa disponibilizasse algumas funcionalidades, nomeadamente:

#### 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.

#### 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.



## DESCRIÇÃO DE 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.

### 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.

## CONCLUSÃO

Achei este trabalho muito interessante, pois não só me permitiu conhecer melhor a linguagem de programação Python, como também novas potencialidades da programação, por exemplo extracção de dados directamente de páginas html.

Achei que o trabalho era algo trabalhoso, sendo necessário gastar muitas horas para o desenvolver.

Penso que os conhecimentos adquiridos durante este trabalho poderão facilmente ser utilizados em futuros desenvolvimentos, quer no meio académico quer no meio profissional.



## BIBLIOGRAFIA

- [i] VelocityReviews, Kill a Thread in Python, Abril de 2004  
<http://www.velocityreviews.com/forums/t330554-kill-a-thread-in-python.html>
- [ii] Python Documentation, Abril de 2011  
<http://www.python.org/doc/>
- [iii] StackOverflow, Abril de 2011  
<http://stackoverflow.com/>

## ANEXOS

### INTERFACE

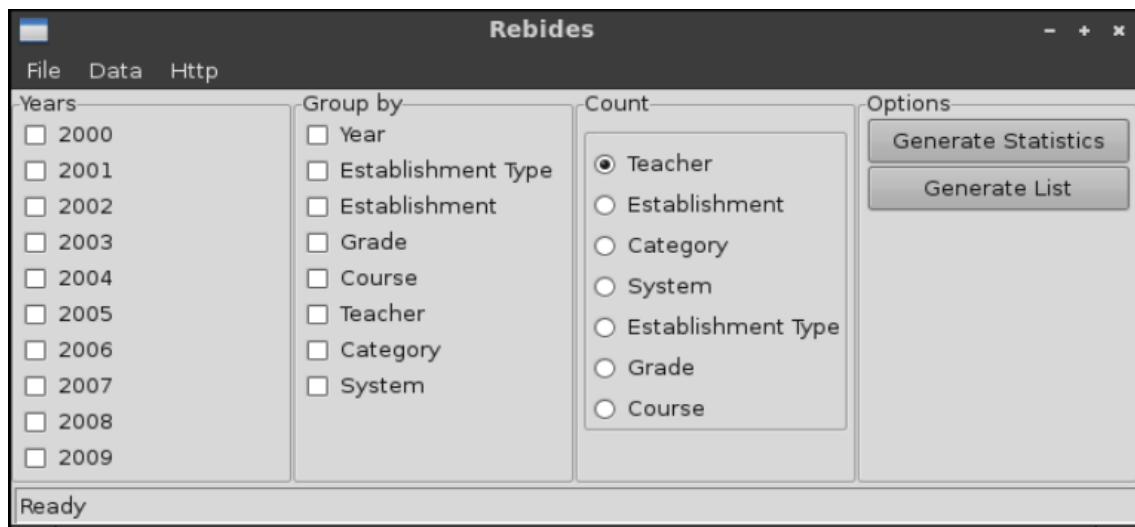


Ilustração 1 - screenshot da interface do programa

### HTML

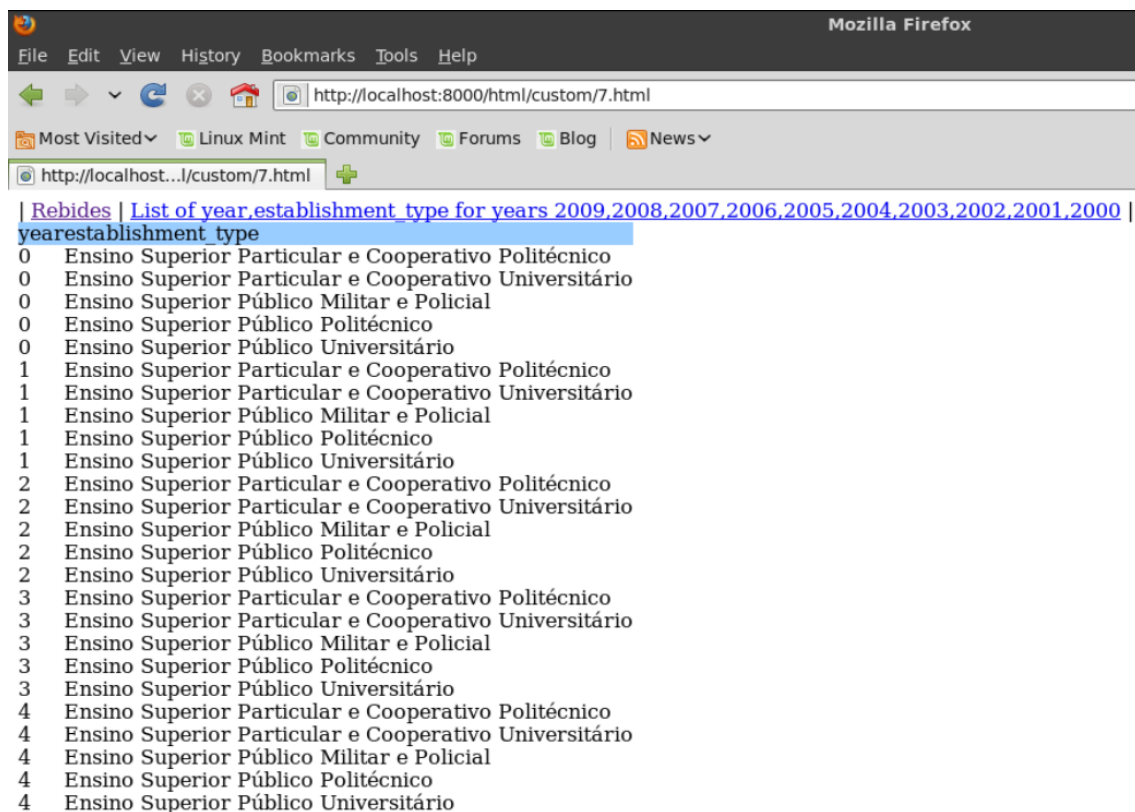


Ilustração 2 - página HTML gerada pelo programa

### CSV



```

|-----
"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

```

Ilustração 3 - ficheiro CSV criado pelo programa

## GRÁFICOS

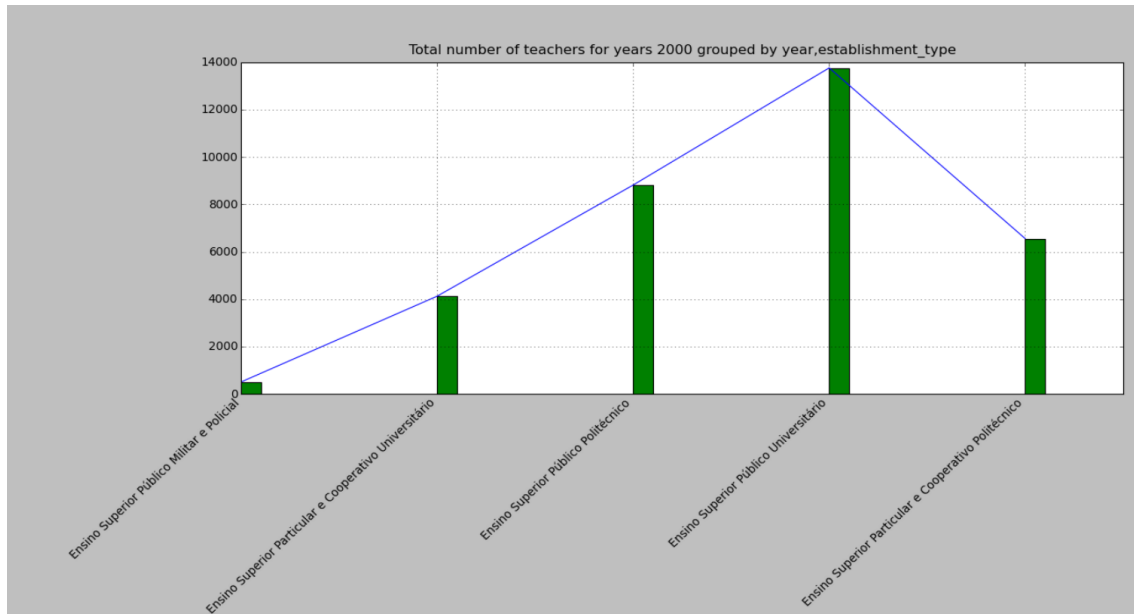


Ilustração 4 - Número de professores por instituição no ano 2000