QUBE Overview bot

In order to get an overview of how the different Process Areas in QUBE is linked, a web crawler has been created to automatically read all the subpages of QUBE, while scanning for the relevant information.

This document outlines the install, usage and settings of this spider bot.

Contents

Install

The bot is written in Python2.7, and is using a library called Scrapy for the web crawling and a library called pydot for creating the graph.

The install part requires local admin rights!

Python and Scrapy

- Install the latest Python 2.7 from https://www.python.org/downloads/ (At time of writing 2.17.12)

  You need to adjust `PATH` environment variable to include paths to the Python executable and additional scripts. The following paths need to be added to `PATH`:

  ```
  C:\Python27\;C:\Python27\Scripts\;
  ```

  To update the `PATH` open a Command prompt and run: (A command prompt can be opened by pressing “WinKey + R”, then input “cmd” and press enter)

  ```
  c:\python27\python.exe c:\python27\tools\scripts\win_add2path.py
  ```

Close the command prompt window and reopen it so changes take effect, run the following command and check it shows the expected Python version:

```
python --version
```

- Install *pywin32* from http://sourceforge.net/projects/pywin32/

Be sure you download the architecture that matches your system (amd64)

Now open a Command prompt to check `pip` is installed correctly:

```
pip --version
```

- If `pip` is not installed, install pip from https://pip.pypa.io/en/latest/installing/
- At this point Python 2.7 and `pip` package manager must be working, let's install Scrapy:

```
pip install Scrapy
```

- Install requests-ntlm, in order to be able to authorize the user login at QUBE:

```
pip install requests-ntlm
```

At this point the bot is able to get the information from QUBE, but it is still unable to produce the expected graph output. In order to produce this, a library called GraphViz is required, as well as a Python wrapper for this library. In the next section, these two are installed.

GraphViz and pydot

- Download and install GraphViz from http://www.graphviz.org/Download_windows.php. (Get the .msi file and right-click the installer, to "Run as Administrator")

You need to adjust `PATH` environment variable to include paths to the GraphViz executable and additional scripts. The following paths need to be added to `PATH`:

```
C:\Program Files (x86)\Graphviz2.38\bin;
```

To update the `PATH` run the "set_path.bat" script, located in the same folder as this readme document. **ONLY DO THIS ONCE!**

- Install pyparsing & pydot:

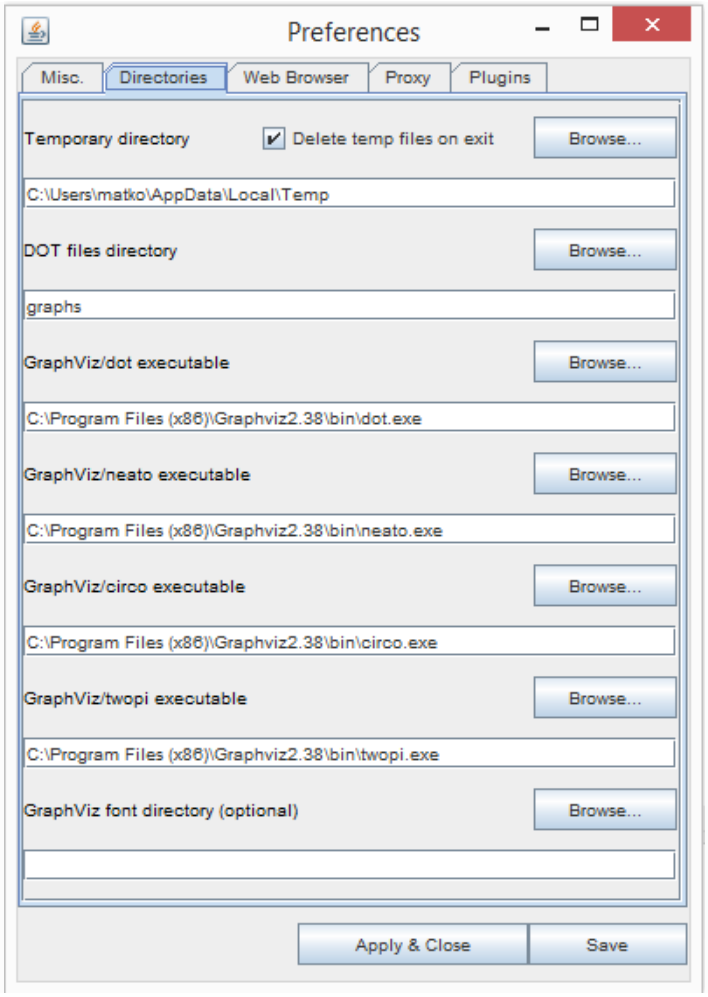
```
pip install pyparsing==1.5.7
pip install pydot==1.0.28
```

That is it for the required install. The next section contains information on how to install an optional viewer, which makes it a bit easier to handle the very large graphs on a computer screen.

ZGRViewer (Optional)

This section will show how to install an optional viewer, which makes it a bit easier to handle the very large graphs on a computer screen.

- Download the ZGRViewer archive from https://sourceforge.net/projects/zvtm/files/zgrviewer/0.10.0/zgrviewer-0.10.0.zip/download.
- Extract the archive somewhere you find soothing.
- Open the extracted folder, and execute the “run.bat” script file by double-clicking it.
- All you need to do now, is to set some preferences in order to let the viewer know where GraphViz is installed. To do this, open “View” and then “Preferences”.
- Choose the “Directories” tab, and fill in all of the GraphViz paths, either by writing the correct paths, or by using the “browse” button. They should look something like this:



- Click “Apply & Close”

Usage

Running the spider

To run the spider in order to create a new, updated graph of QUBE, the bat script in the "QUBE_Walker" directory can be used. The spider is started by double-clicking the "run.bat" file. The spider takes about half an hour to crawl the entire QUBE site, and put together a graph.

NOTE: The first run, you need to enter a QUBE username and password in the settings file. See the Settings section for details on how to do this.

Once the program is done, the finished graph can be found in the "graphs" subfolder, named in accordance to the naming settings given (Default by date and timestamp).

Viewing the PDF

The PDF files produced by the program, is too large to be viewed in Adobe Reader, but can be opened in any other PDF viewer, e.g. Power PDF Advanced, which is available in OneWebshop. (Please note that once it is opened, you need to zoom to be able to see anything).

Using ZGRViewer

The optional viewer for .svg files makes it easier to manage and view the large graphs on a computer screen. The viewer is installed as per the install instructions above, and is started by running the "run.bat" script in the folder. Once opened, the QUBE graph can be loaded by clicking "File", "Open" and the choose "Open SVG generated by GraphViz…". Browse to the graph output folder created in the QUBE_Walker directory, and choose the desired graph with the ".svg" extension.

The viewer contains a number of tools, some of which is more useful than others. They can be found by moving the mouse to the left edge of the viewer in order to reveal a small menu as seen below.

Furthermore the viewer allows searching in the graph by clicking "View", "Find..." or CTRL+F.

Settings

The settings of the program can be found in "**settings.py**". There should be plenty of comments, describing the different options, so only the most important will be highlighted here.

- **Username and password for QUBE: (initials on XXXXX and Dong Energy password) [REQUIRED]**

```
QUBE_USER = 'DE-PROD\\XXXXX'
QUBE_PASS = ''
```

- **Settings defining the separation of the elements in the graph**

```
DOT_MARGIN = '0.15'          # Margin around text inside figure
DOT_RANKSEP = '5'            # Vertical separation of levels in graph
DOT_NODESEP = '0.9'          # Horizontal separation of nodes in same level
```

- **The number of graphs to save, before starting to automatically delete the oldest**

```
NUMBER_OF_GRAPHS_TO_SAVE = 10        # (-1 = infinite)
```

- **Whether or not to save the graph as a PDF files as well as SVG and DOT**

```
SAVE_PDF = True              # Export a PDF file of the structure
```

- **Which top level areas to include in the graph (Must match the exact names in QUBE)**

```
QUBE_AREAS = ['Governance and Compliance', 'Core processes', 'Support Processes']
```