

Documentation for Database Acquisition Program

Greg Knoll - iEXERGY 2014

Structure:

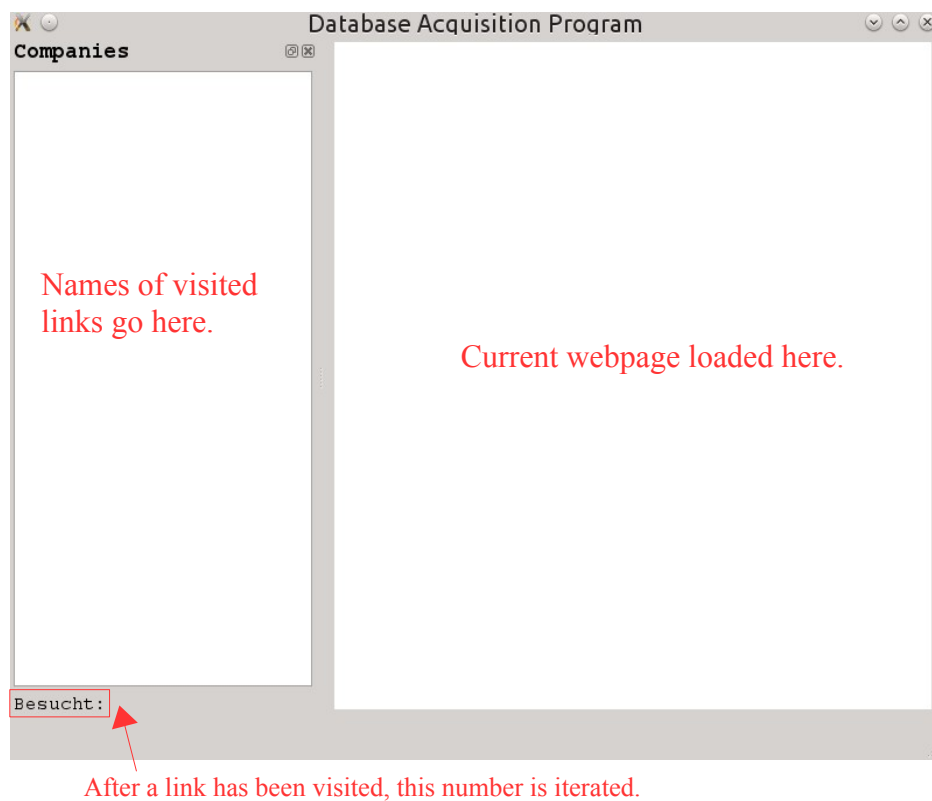
```
main.cpp
window.cpp/.h
innerwindow.cpp/.h
```

How it works:

1) Enter **URL** of website into main.cpp, line 49.

```
49      window.setUrl(QUrl("http://www.hwk-do.de/index.php?id=727"));
```

2) This creates a new UI window (window.cpp) as shown below:



3) The webpage is traversed, and each link (<a> tag) that contains the word "betrieb" is visited. The name of the company contained in the link is added to the Companies margin.

4) When each link is visited (innerwindow.cpp), the DOM or document structure is traversed. This structure needs to be adapted for each website. When the desired element has been reached (i.e. a table row <tr> element), the toPlainText() is then called on this element and its information is sent to the output file.

Adaptations for each web site:

- 1) main.cpp, line 49 : Change the URL in window.setURL();
- 2) window.cpp, examineChildElements(), line 95:
Change the name contained in the links to individual pages.
In this example, "betrieb" is the keyword used.
- 3) window.cpp, examineChildElements(), line 103:
Change name of button to move to next page.
In this example, "weiter" is the name of the button.
- 4) window.cpp, parse_page(), line 122:
If link from examineChildElements() is not a complete URL, then
need to add homepage info. Otherwise, appropriately modify the
URL before passing to InnerWindow constructor.
- 5) innerwindow.cpp, get_biz_info():
Change DOM structure.