**Using the WebFileExtractor**

1. Open up the example web file extractor implementation file (WebFileExtratorImplementaton.py).
2. Adjust your desired settings  
   1. Domain : the automatically crawls through the website collecting links. It clicks on and gathers all links in pages starting with this link. Eg if this is set to [www.mypage.com](http://www.mypage.com), any link starting with mypage.com will be clicked and links in that page will be added to the list.
   2. url : the program will start on this page
   3. filetype : this is the file type/name you would like to download. Eg for pdfs type “.pdf”. Any links ending in .pdf will be downloaded. If you require all videos/images or audio files, you can use the following flags “—IMAGE”, “—VIDEO”, “—AUDIO” instead of a file type.
3. Run the program

**NOTE:** the web file extractor will save all the files in Web\_File\_Extractor/downloads, a text file will also be saved to record files that have already been downloaded. Each link is stored as the address and the filename it was saved as. (Replacing / with {WS\_FSLASH})