ASSIGNMENT 5.3

Design a crawler in C++ which will take a query as input and download the result links provided by Google in response to that query. The crawler needs to have the following features.

1. The searches and downloads will be through any of the following proxy servers
   - 10.3.100.211 8080
   - 10.3.100.212 8080
   - 144.16.192.216 7777

2. Same link structures as the original sites have to be maintained locally.

3. Following restrictions can be imposed.
   - Download for each result link can be restricted by size.
   - For each link, the download can be restricted to pre-specified depth and breadth.
   - Download for a query can be restricted to first N result links returned by Google.
   - It can download files with any specified extensions.
   - It can download images with any specified extensions.
   -

   The restrictions have to be read from a setting file.

4. Multithreaded download is encouraged.

5. Render the results for a query through an html page.

NOTE: Following APIs can be used for this assignment.

- For connection with Google
  i. API: Use **gSOAP 2.7.6e stable**. It can be downloaded from http://sourceforge.net/project/showfiles.php?group_id=52781&package_id=68161 .

- For downloading individual URL
  i. Use **curl 7.15.3**: It is available at http://curl.haxx.se/download.html.