Web Crawler

Ankit Nimje

Surbhi Kanthed

In this week we managed to develop our web page by integration back end program with local server. Two divisions were created each for links and images. Left dialog box is assigned with links which will give all crawled web page links and right dialog box will provide all related crawled images. Image dialog box gets a scrollbar if needed.

While executing python script, we got the links on local server in JSON format. Then in our html file, we used AJAX to call python script at back-end and stored the JSON result in object variable, then created anchor object and converted all string object type of JSON into 'a href' link of HTML. These all links have been passed to container of dialog boxes which we can see now.

To display all images of web pages, first we tried to fetch all the links of the images in a particular web page and then we stored these images in a local folder and tried to import that folder to display images in our dialog box. But somehow this technique did not work. Then we fetched all the respective links of the images and used AJAX to run python script and get all links of JSON format and convert them into a image object. With the help of this we were directly able to fetch all images when HTML page runs or user search a particular input.

Problem Faced

References:

1] S. Chakrabarti, M. Berg, and B. Dom," Focused crawling: a new approach to topic specific Web resource discovery", Computer Networks, vol. 31, pp. 1623-1640, 1999.

2] W3resource.com – Python Flask

3] Tutorialspoint.com – Python Flask