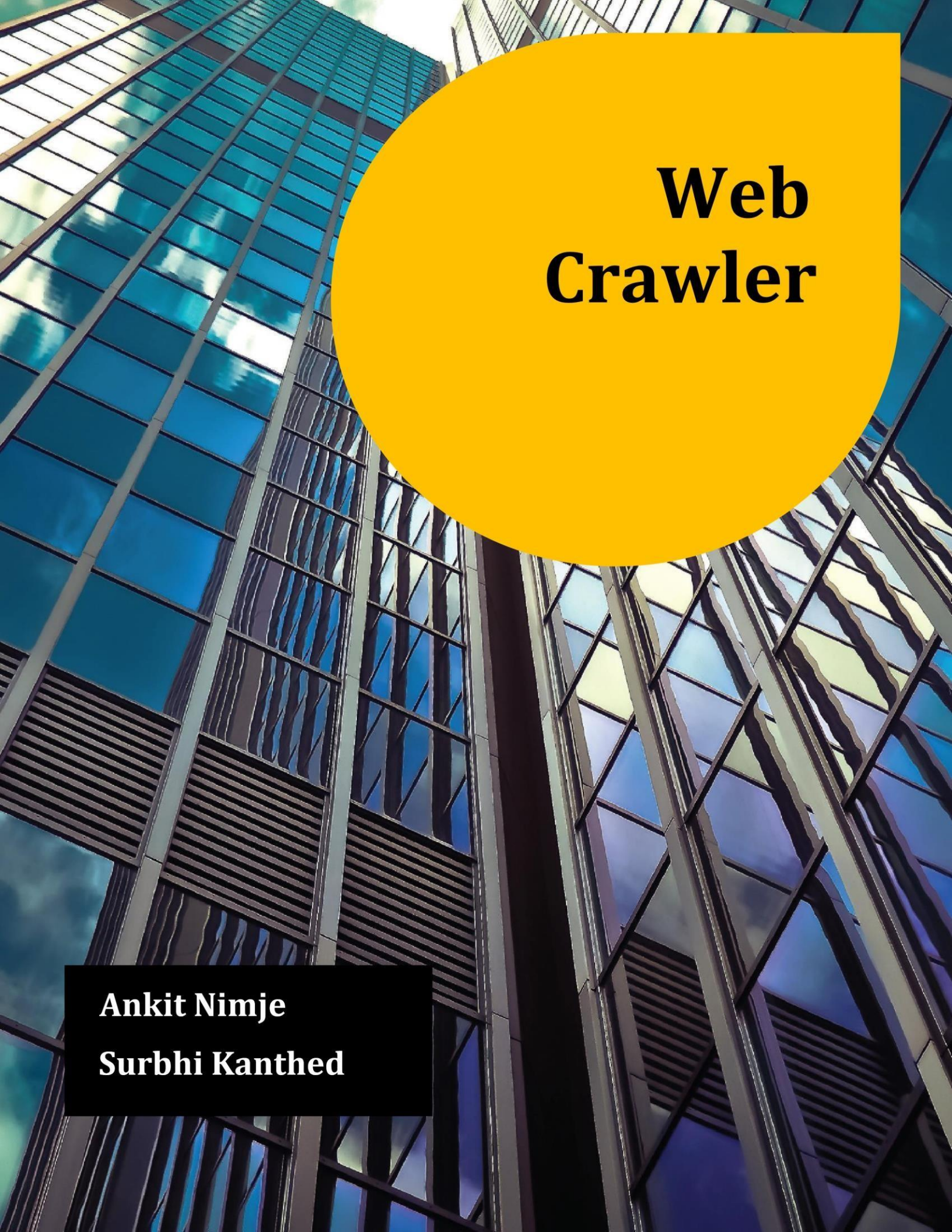

Web Crawler

Ankit Nimje

Surbhi Kanthed

In this week, we were working mostly on back-end programming. Until very recently, we were fetching all webpage links in root folder but now, all those links are directed to localhost. Even though it looks more disorganized, we plan to work on it by coming submission.

Also, we are also working to fetch all the images in our crawled pages similar to Google Image search. We managed to fetch all the images from selected web page but plan to integrate this feature in our crawler program. In the end you should able to see result web page with two divisions. First division will provide crawled page indexes and second division will be for image search.

Moreover, we managed to obtain query text from home page which will be later provided to crawler program. To cross check, passed value is printed in a dialog box in json format.

Problems Faced:

Initially it was difficult for us to link python to HTML as input is provided to python program from HTML page and get back the results to HTML page. After some efforts we figured out, it can be done using Python Flask, a web framework written in python and still working on it.

Implementation:

- Index.html ---front-end interface.
- WikiImageUrl.py --- To fetch all the images from particular website
- Crawl_data/app.py --- Crawled web links presented on localhost

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