

A low-angle, upward-looking shot of a modern skyscraper with a glass facade. The building's structure is composed of dark metal frames and large glass panels that reflect the sky and clouds. A large, bright yellow circle is positioned in the upper right quadrant of the image, partially overlapping the building's facade.

# **Web Crawler**

**Ankit Nimje**

**Surbhi Kanthed**

To implement a good Search Engine, rather than presenting webpages randomly with our keywords, it is better to sort those webpages according to occurrence of query words. In this week, we studied how to design a focused web crawler while keeping in mind a plan to present them in ranked manner.

As like a Search Engine, which finds webpages with passed query, our crawler is able to fetch webpages with passed keyword. For this purpose, we designed a classifier and distiller which checks for occurrence of keyword in particular webpage and presents those web pages. In coming week, we will try to rank those results according to page rank (according to occurrences of keyword). Also we plan to crawl through those webpages with high page ranks as those results will lead us towards better content related to our keyword.

### **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] M. Najork and J. Wiener," Breadth-first search crawling yields high-quality pages", in Proc. 10thInt. World Wide Web Conf., pp. 114-118, 2001.