Web Scraping: How Companies Can Use ‘Big-Data’ generated by Facebook?

Somdip Dey, Juncheng Li, Shaohsuan Cheng
School of Computer Science,
The University of Manchester.
United Kingdom.

Introduction
In the era of social networking websites and technology advancement, the companies which know how to make sense of the data generated by OSN users and use them properly, hold a beneficial position in the market. Analyzing the data generated by OSN users can provide loads of information that can be used by the companies for various purposes such as customer feedback, knowledge of customer requirement, customer’s preferences, etc.

Two important tools used by data aggregators to collect data from websites are web crawling and web scraping.

Motivation

Every day Facebook ingests more than 500 terabytes of data, making it a silo of data, but most of the companies are unarmed with a systematic way of extracting data for their reuse or benefit.

Two Way Approach

Type Of Extractable Data on Facebook

- Profile Data
- The Social Graph
- Traffic Data

Data Extraction Methods

- Public Listings
- False Profiles
- Profile Compromise and Phishing
- Malicious Applications
- Facebook Query Language (FQL)

Web Scraping and Crawling to extract data from Public Listings, and FQL to extract data from Social Graph.

Conclusion
Using the proper web scraping or crawling tools along with FQL will give the companies a methodical and systematic way to collect essential information from the enormous amount of data generated by OSN users.

References