Abstract

Increased amount of internet users leads to collision in the names on the web forums and in many scenarios, which in turn leads to maximum number of users who are using their aliases in the web. This creates a difficulty in detecting the proper user. So, systems are suggested to identify their aliases using the entity graphs. But most of them are experimenting on the datasets; on the other hand very few systems exist to be worked on real entity. So implemented system put forward an idea of finding aliases on the real web data by using an enhanced web crawler which collects all sub URL’s of the given seed URL, which is analyzed by the another baby crawler to fetch and parse the web data as human readable content using random walk relational theory. Alias graph is identified to be more efficient with the help of real relation entity graph on the collected web data.

References


Index Terms

Computer Science

Algorithms

Keywords

Web crawler, NLP, Entity relation, Random walk, Cauchy distribution.