Abstract

The modern developments through the internet have the opportunities to access heterogeneous information system anywhere in the world. We can integrate information which is structured or unstructured and deliver it to any system on big data platform. The information may be from heterogeneous sources and with different representations. In this paper, we propose an integration scheme for books data from different sources over the web.

Graph database allow simple and rapid retrieval of complex hierarchical structures that are difficult to model in relational systems. Nowadays graphs have become very popular in various domains like social media analytics, healthcare, chemistry, business intelligence, networking and many more.

We have proposed a system that integrates data of books from various sources across the web. Here graph database is used named TITAN. We can easily retrieve complex information using simple queries using Gremlin, a graph query language.
References


22. URL project, http://cURL.haxx.se/
23. Python re Module, https://docs.python.org/2/library/re.html

Index Terms
Computer Science       Databases

Keywords
Heterogeneous sources, graph database, information integration, graph query