Information Integration from Heterogeneous Web Sources

Shahid Kamal, Dr. Roliana Ibrahim
FSKSM, Universiti Teknologi Malaysia
skamaltipu@gmail.com, roliana@utm.my

Abstract

Information integration is the problem of providing a unified and transparent view to a collection of data stored in multiple, autonomous and heterogeneous data sources. In this study we are focusing only on the inter organization information integration (web), as to different web sources have different data base management systems for the data management and different DBMS’s has their own syntax and semantics. The previous studies had identified several problems in this area like automatic source wrapping, construction of global schema, mappings between sources and global schema etc; integrating information from heterogeneous sources associates to different issues like syntactic and semantic. This study is focused that the mediator based information integration model could be improved by implementing the selected technique to overcome the semantic issues related with heterogeneous information integration from web sources. In this study we will concentrate on the querying problems related in mappings to increase the semantic interoperability.

Keywords: Information integration, heterogeneous data sources, syntactic issues, Patternbase technique