

Chapter 12

Proposition of a New Ontology–Based P2P System for Semantic Integration of Heterogeneous Data Sources

Naïma Souâd Ougouti

Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf, Algeria

Hafida Belbachir

Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf, Algeria

Youssef Amghar

INSA Lyon, France

ABSTRACT

Semantic web offers new opportunities to multi-sources integration field, and many approaches like P2P systems are revisited taking into account the new requirements. In this chapter, the authors present their P2P heterogeneous and distributed data integration system. It is a super-peer system, where peers are regrouped by type of data (relational, image, text, etc.) around a super-peer which contains a domain ontology. Peers data sources are exported in a common format in the form of a semantically rich ontology. Schemas reconciliation is done by matching domain and local ontologies by the use of a similarity function whose contribution is based on the direct and indirect semantic neighborhood. Queries are described using ontologies, then routed towards relevant peers thanks to a semantic topology built on top of the existing physical one.

INTRODUCTION

Exponential development of information exchange through Web has increased the difficulties to find relevant information which is represented and stored in a multitude of heterogeneous data sources. Several solutions were proposed and are rather old. We have seen the emergence of data warehouses, mediators and P2P systems.

DOI: 10.4018/978-1-5225-5384-7.ch012

Proposition of a New Ontology-Based P2P System

The data warehouse approach consists in carrying out integration by building real databases gathering relevant information of considered applications. The user will work directly on the data stored in the warehouse.

The mediator approach is a method where the data is accessible only from the information sources, the user in this case will work on abstract views built with the aim of describing the various data sources. Searching information from such sources requires the construction of execution plans to obtain the whole results from information sources.

The Peer-to-peer (P2P) solution is a recent approach; it can be seen like a generalization of mediators/data warehouses architectures. These integration systems follow a decentralized approach for integration of autonomous and distributed peers containing data which can be shared. The principal objective of such systems is to provide a semantic interoperability between several sources with the absence of global schema (Ougouti, Belbachir, Amghar, Benharkat, 2010).

With the advent of semantic Web, new possibilities are offered and many traditional approaches are revisited giving by these interesting results. We can say today that semantic Web and P2P systems have opened up several research possibilities in data integration field that includes a variety of knowledge-based techniques, like semantic data modeling, ontology definition, query translation, query optimization, and terminology mapping. The main objective of semantic Web is to add semantics to Web data sources and allow data from diverse sources (possibly stored using different schemas) to be accessed seamlessly. This is possible by the use of multiple tools such as Ontologies which provide a vocabulary that describes a domain of interest and a specification of the meaning of terms used in the vocabulary (Euzenat & Shvaiko, 2016).

We have noted that the majority of integration systems treat a maximum one data model or two at the same time and do not allow complex and multimodal queries whose results can be various types of data like texts, videos and images. The goal of authors is to propose solutions to these problems by presenting in this article, the new version of their data integration system in a P2P environment called MedPeer. In the first version of this work (Ougouti, Belbachir, Amghar, Benharkat, 2011), they presented a very basic architecture that does not reflect the actual one. There is a great difference between the content of the peer and super-peer of the two versions. They will give more details on new functionalities, present mediation process that relies on a new similarity function, Query description with an ontology, semantic topology construction and semantic routing solution.

This paper is organized as follows:

In the next section, the authors present a state of the art of the most representative approaches of integrating data in a P2P environment. The new version of MedPeer Architecture is introduced by presenting in detail the peer and Super-Peer structure and their different functionalities. Then, the source description module which is an important component of a peer is taken up. The authors describe the solution proposed for calculating similarity between domain ontology concepts and local ontologies ones in the main function of the matching manager module. Subsequently, they explain how the semantic topology is built and present modules related to the query management: query description with an ontology, Relevant Peers search module and Query reformulation module. The main results of experiments for evaluating the similarity function and query semantic routing are explained. Finally, the authors complete this chapter by future research directions and conclusion sections.

29 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/proposition-of-a-new-ontology-based-p2p-system-for-semantic-integration-of-heterogeneous-data-sources/203427

Related Content

Evaluating e-Learning Initiatives: A Literature Review on Methods and Research Frameworks

Stelios Daskalakis and Nikolaos Tselios (2013). *Web-Based and Blended Educational Tools and Innovations* (pp. 163-179).

www.irma-international.org/chapter/evaluating-learning-initiatives/68644

A Novel Method to Dynamically Fix Threshold for Node Neighbourhood Based Link Prediction Techniques

Anand Kumar Gupta and Neetu Sardana (2020). *International Journal of Information Technology and Web Engineering* (pp. 17-34).

www.irma-international.org/article/a-novel-method-to-dynamically-fix-threshold-for-node-neighbourhood-based-link-prediction-techniques/241774

PECA: Power Efficient Clustering Algorithm for Wireless Sensor Networks

Maytham Safar, Hasan Al-Hamadi and Dariush Ebrahimi (2013). *Network and Communication Technology Innovations for Web and IT Advancement* (pp. 239-248).

www.irma-international.org/chapter/peca-power-efficient-clustering-algorithm/72765

WDPMA: An MA-Based Model for Web Documents Prioritization

Santosh Kumar and Ravi Kumar (2021). *International Journal of Information Technology and Web Engineering* (pp. 1-24).

www.irma-international.org/article/wdpma/275731

Social Network User Profiling With Multilayer Semantic Modeling Using Ego Network

Tamil Selvi P., Kishore Balasubramaniam, Vidhya S., Jayapandian N., Ramya K., Poongodi M., Mounir Hamdi and Godwin Brown Tunze (2022). *International Journal of Information Technology and Web Engineering* (pp. 1-14).

www.irma-international.org/article/social-network-user-profiling-with-multilayer-semantic-modeling-using-ego-network/304049