



MULTIPLE VIEWS IN PEER DATA MANAGEMENT FOR E-COMMERCE APPLICATIONS

M.B.AI-Mourad ¹⁾, Rozalina Mohamed ²⁾, Yaser M. A. Khalifa ³⁾

¹⁾ College of Information Technology
University of Dubai
Dubai, UAE

²⁾ School of Engineering & Applied Science
Aston University
B4 7ET, UK

³⁾ Electrical & Computer Engineering,
State University of New York
New Paltz, NY 12561, USA

Abstract: *This document presents the required layout of papers to be submitted for publication in the "Computing" International journal. The abstract may not be longer than 150 words. Peer-to-peer (P2P) systems are revival paradigm for information sharing among distributed nodes in the network. A P2P network is a network that relies primarily on the computing power and bandwidth of the participants in the network rather than concentrating it in a relatively low number of servers. P2P software systems like Kazaa and Napster rank amongst the most popular software applications ever. Numerous web businesses and sites have promoted "peer to peer" technology as the future of Internet networking for E-commerce. Multiple views for data are created for mediating between data sources on the Semantic Web. Our goal is to support users' different needs. This is due to the fact that different users have different needs for joining the P2P community and their requirements may change over time as new information become available. Hence the same information may participate in many different ways in multiple data sources' mapping efforts.*

The idea of creating multiple data views has been adopted in federated database systems. However, it is not yet implemented in peer-to-peer systems. This is due to the lack of shared knowledge among peers. Moreover, it is not trivial task to create multiple data views in an unpredictable environment such as P2P.

This work investigates the possibility to construct multiple data views in P2P environment. This will be achieved through the development of a tourism application framework. The described framework would benefit small businesses by participating in the P2P network without being forced to use compatible standard indexable-web data-bases provided by large corporations.

Keywords: *Peer-to-Peer Data Management, Peer-to-Peer data integration system, Multiple data views, Database Integration, E-commerce.*

1. INTRODUCTION

The World Wide Web (WWW) and search engines have dramatically succeeded in providing an easy way for users to share and distribute data. Most users turn to WWW search engines when looking for specific information. But a recent survey of the 11 most widely used search engines reported by NEC Research Institute shows the fact that no single search engine covered more than 16 percents of the Web's contents [26]. Besides the size of search engine's index, searching quality is comprised of a variety of factors including page-depth amount (index the full or partial text on the Web), updating the index and finding the most relevant indexes [34].

First, let's see how some search engines operate? Search engines may collect all related data based on a given keyword and return the result to the user's desktop. Searching is done by the collection of the indexable-web that has been stored previously. This means that the search engine is restricted to supporting keyword-based search only within its index. In this case, how could the WWW search engine ensure the results that has been returned to user's desktop are updated and the most relevant to user's needs? Recent trends in research suggest that we may ultimately move to a more extreme peer-based model and semantic web in order to improve the searching precision [9].

Peer-to-peer (P2P), Figure 1., systems are revival paradigm for information sharing among distributed nodes in the network. A P2P network is a network that relies primarily on the computing power and bandwidth of the participants in the network rather than concentrating it in a relatively low number of servers. P2P networks are typically used for connecting nodes via largely *ad hoc* connections. Such networks are useful for many purposes. Sharing content files containing audio, video, data or anything in digital format is very common, and realtime data, such as telephony traffic, is also passed using P2P technology. P2P software systems like Kazaa and Napster rank amongst the most popular software applications ever. Numerous web businesses and sites have promoted "peer to peer" technology as the future of Internet networking for E-commerce. This is true for small businesses, and home-base businesses in particular, P2P would be most useful. P2P would provide networking in addition to resources sharing among multiple nodes of small business. It would save large investments in servers' purchases.

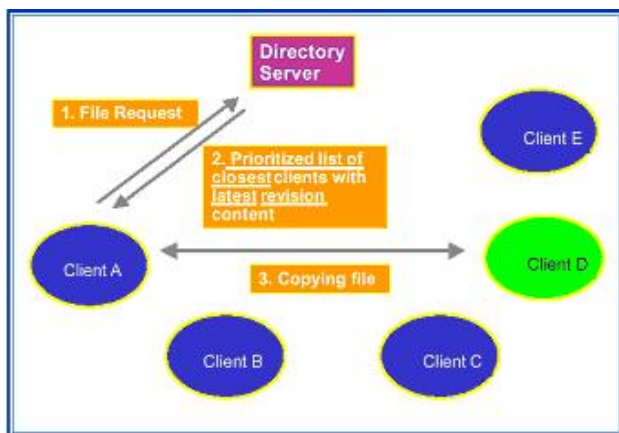


Fig. 1 – A typical P2P Network

The Semantic Web is a web of data. There is lots of data we all use every day, and its not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar? The World Wide Web Consortium (W3C) expects that the semantic web is capable of enhancing the machine-understandable semantic of information that is being processed [3]. Therefore, semantic web will be able to help improve searching precision because it is able to process machine-understandable semantics of information. From a technical point of view, P2P are particularly well suited for networked host to share resource in distributed manner and information searching [11]. P2P is a suitable solution because it makes possible for different participants (peers) to

maintain their own knowledge structure while exchanging information. At the same time P2P are potential architecture to support knowledge-based search (searching that is not limited to keyword-based), where semantic of information can be taken into account.

In this paper we propose a novel way to construct a multiple views for peers who join the Peer Data Management Systems (known as PDMS). The multiple views are created for mediating between data sources on the Semantic Web. Our goal is to support users' different needs. This is due to the fact that different users have different needs for joining the P2P community and their requirements may change over time as new information become available. Hence the same information may participate in many different ways in multiple data sources' mapping efforts.

The rest of this paper is organized as follows: Section 2 provides a review of the related work in supporting view generation. Section 3 introduces a proposed framework for an E-commerce application in which tourism is used as a practical paradigm for further study. Section 4 discusses the framework general features and some related work on PDMS. Section 5 identifies a set of open research questions on implementing multiple views. Discussion and Conclusion come in final section.

2. RELATED WORK

A framework for dealing with multiple view integration is presented in [8] [19] [20]. Basically, the view integration is divided into several main processes: (1) Obtain a required keyword from user. In order to avoid missing relevant result, we need to expand the search by getting some synonyms from ontology. (2) Combine a set of schemas (or views) by defining a new schema that contains all original views, plus a new set of interrelational dependencies that express how data in distinct views is interrelated. (3) Schema learning applies to integrated schema (schema or views that has been combined previously). It is needed to obtain rules for schema or view optimization process. (4) Integrated schema or view is optimized by reducing some redundancy among schemas and minimizing the integrated schema size as in database normalization process.

A knowledge base structure to support multiple view generation is presented by Karunaratna et.al [23]. The knowledge base is built by extracting and merging metadata of source databases. Al-Mourad et.al has proposed a semantic materialization rules to support multiple view generator [2]. The rules are capable of directing the global query processor to combine data from different databases and reconcile

database heterogeneity. Since the focus of these works is for federated database, knowledge is elicited from the database owner as they join the federation and when they evolve the databases. Due to the nature of P2P network where federated database and global knowledge is unavailable, deriving source meta-data in P2P environment is not a trivial task.

With the growing popularity of XML, system that learns mappings between XML source schemas and integrated schema called LSD is presented by Doan et al [10]. Given an integrated schema, first set of data sources have been manually mapped to the integrated schema. Then the system learns from these mappings to propose mappings for new data sources. Integrated schema in LSD is given by human, while Jeong and Hsu [20] has generated it dynamically and automatically. The main contribution of their work is on mappings between different XML DTDs. A tree grammar-based approach for inducing integrated views for XML data with heterogeneous DTDs are converted into a tree automaton. States belonging to similar types are merged to obtain a minimized integrated view. Integrated view and source description are automatically derived by a view inference system. However, the derivation of both works is conducted offline. This means that they are not considering a dynamic state of network which is unsuitable for P2P network where the environment is dynamic.

In this research we consider how dynamic environment can emerge by adding and deleting links between peers iteratively according to the evolving knowledge about other peers in the network. Therefore, we should fully utilize local knowledge of peers that is adaptive to its environment.

3. PROPOSED FRAMEWORK

We describe the need of multiple data views in P2P environment by using a scenario of tourism, as presented in Figure 2. Currently, major tourism websites such as Expedia, Price-line and others have a large network of providers for Hotels, Flights, etc. However, small business providers of tourism services have to set-up their databases in a standard searchable indexed fashion dictated by the Web Search Host. For small businesses, this adaptation process could be quite costly, and unbearable financially.

The scenario consists of several information providers and customer. Each information provider has different roles. Each of them would advertise their own services and may use their own terms. Let's consider four main parts of information provider that form the basic tourism services:

ACCOMMODATION, TRANSPORT, TOURISM AGENCIES and TOURIST ATTRACTION AREA. ACCOMMODATION consist of company whose offer the accommodation services, such as hotel, resort and etc. TRANSPORT is a group for agencies whose offer the transportation services. TOURIST ATTRACTION AREA is a collection of places of interest. This is may be used by tourist information centre or anybody who want to advertise the attraction places. The TOURISM AGENCIES is a group for those who provide the travelling services such as travelling agencies and tour operators. CUSTOMER is who ever inquire this information.

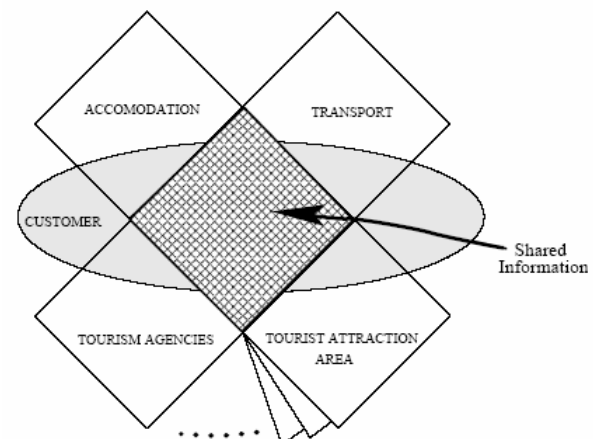


Fig. 2 – Shared Information Between Different Roles in Tourism Scenario

With the WWW search engines, if a customer wants to get the travel information about 'New Paltz' they may use the word 'New Paltz + travel information' as a keyword. Luckily, they may get some information regarding to attraction places, transportation and hotels whose post their information to that the portal that they have accessed. Today we can use some web pages or web portals that provides travel information guide. By using the portals, we are able to get information on accommodation, transportation advice and list of attraction places as well. Unfortunately this kind of information exchange that currently occurs among different information providers or between information providers and web portal is time consuming, out-of-date and error-prone, although it is often available in electronic form [11]. Customer is accessing any web page or web portal whose work on a central integrated data sources with a single view (mediation between user and integrated data sources). We cannot assume that whoever accessed to the 'same' data should get the same answer. For example, the result should be different between customers from different geographical location. It means, result for customer in 'London' should be different with customer whose in 'New Paltz' itself. Some differences such as, currency used when

describing the prices, information on how to get to specific places and so on. Therefore multiple views are essential as mediation between customers who send the query and multiple data sources provided by information provider.

Multiple view approaches have been implemented in federated database [18] [27] [29]. However the same approach seems not to work in P2P environment due to the fact that central integrated data or central knowledge about data sources is unavailable. Multiple views research in federated database eliciting knowledge from database owner who join the federation [2] [1] [22] [23]. Participant in P2P community is not as-fix-as in federated database, where each peer (participant) may join and leave the community as they like. The only knowledge about other peers in the network is obtained through local knowledge of peers. In P2P local providers who might be too small to participate in a large www search engine, and who in the same time could have inside-information about the local destination, could find themselves now competing on customers logging on to the P2P network.

4. PEER DATA MANAGEMENT SYSTEMS

Schema-based P2P systems, also called Peer Data Management Systems (PDMS), are a natural extension of federated database systems [35]. PDMS offer a flexible architecture for decentralized data sharing. In addition to the characteristics evolving from the P2P paradigm, each peer in a PDMS provides its own data with its own schema. Each peer is associated with a schema that represents the peer's domain of interest and semantic relationships between peers are provided locally between pair of peers. In PDMS, each peer exports data in terms of its own schema. Then the schema is being mapped to local schema of peer who's done the export. Peers are autonomous systems and mappings are dynamically created and changed. For that reason, query answering process in PDMS is not a trivial process because schema mapping need to be considered for each query routing and re-routing process [6].

PDMS has no global knowledge, neither a global schema nor information of data distribution or indexes [21]. The only information a participating peer has is about its neighbour. In the absence of centralized mediator with global knowledge that can directly routing the query message to the right peer, finding the right peers for querying for efficient query routing is not a straightforward process [16].

In certain circumstances, data for query result are retrieving from multiple sources. This query involves locating and retrieving relevant data among

one or more data sources. When integrating information from several data sources, the same data (in the real world objects) can exist in multiple different formats and structures. As illustrated in Figure 3, 'NAME' and 'AGENT-NAME' are referred to the same semantic meaning even their annotation is dissimilar. Besides some differences in annotation (tag), there are differences in element structure that refer to the same meaning. 'Atlanta, GA' is a data of attribute 'CITY' that belongs to a sub-child element tag 'DESTINATION' in file 'a' while it is file 'b', it is the data of sub-child element tag 'CITY'. It means that both of them are belongs to different data structure. Detecting object semantic similarity from multiple data source at schema and data level has been discussed by some previous researchers [36] [37] [7] [2]. Efficient object detecting similarity process in P2P systems poses a variety challenges mainly resulting from the limited global knowledge. Particularly with regard to query answering processing, existing approaches cannot be applied easily to schema-based integration in P2P environment.

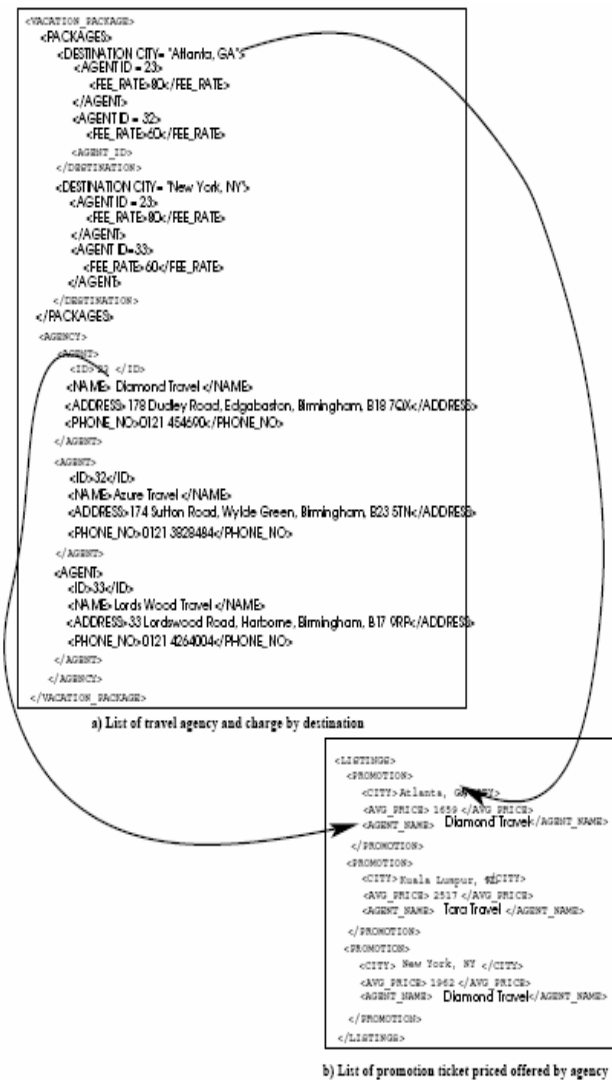


Fig. 3 – XML Files With Different Annotation

The research community has produced several peer data management systems that aim to enable interoperability among distributed heterogeneous data sources. The following subsection is our study of related works focuses on complete systems.

4.1. SWAP

The SemanticWeb and Peer-to-peer (SWAP) [11] project is an ontology-based knowledge management for P2P environment. Knowledge structure is stored on local repository. The knowledge structure is an integrated structure of source between local and other peers. The local repository is shared among peers. The knowledge structure is also a knowledge sources. The knowledge sources are represented as RDF(S) graph. The knowledge sources are integrated automatically into the local repository. Any changes in the knowledge source are propagated to the local repository. However, changes in local repository are not propagated to other peers. RQL-related query language is used in SWAP. If its inference engine cannot get the answer from local repository, it will split the query and distribute the sub-query in the P2P network. This query has to be rewritten in order to fit the knowledge structure on other peers. Returned answers are assembled together to the user without any filtration. Issue on personalizing users' need and mediation is not being mention in SWAP. Whereas, our project focus on constructing data views based on users' preference.

4.2. BIBSTER

Bibster [15] is a P2P system based on the JXTA [13] [12] platform. JXTA technology is a set of open protocols that allow any connected device on the network ranging from cell phones and wireless PDAs to PCs and servers to communicate and collaborate in a P2P manner. Bibster is embedded in SWAP architecture, combining the Peer-to-Peer paradigm with Semantic Web technologies. Bibster focus on the exchange of bibliographic data among researchers issued semantic topology [16] in which peers can advertise their knowledge in the network. Other peers may accept this advertisement, thus creating a semantic topology. Similarity function, 'Sim' is used for determining peers to forward the query and to detect duplication of query result. However, this project does not focus on schema heterogeneity. All schemas for bibliography have been pre-defined, based on two common ontology, i.e ACM or SWRC Topic Hierarchy. This means that schema being shared is not heterogeneous. In contrast to Bibster, our work issues of heterogeneity on schema level, which allows for information sharing with different schemas relying on local mappings between schemas.

4.3. EDUTELLA

The Edutella project [31] [32] [5] is a multi-staged effort to scope, specify, architect and implement a RDFbased metadata infrastructure for P2P-networks based on the JXTA framework. This project focuses on exchanging of learning material. Each peer can make its meta-data information available as a set of RDF statements. It makes the individual RDF peers connected to the Edutella network completely transparent by specifying and implementing a set of Edutella services. The services are query, replication, mapping, mediation and clustering. The mapping services will be able to manage mappings between different schemas and use these mappings to translate queries over one schema to queries over another schema. In Edutella, each peer makes meta-data available as a set of RDF statements. However, today most commercial and scientific applications have facilities for automatically exporting their data into XML form [17]. Hence, RDF data form is still not being available broadly. Edutella may need semantic integration of XML and RDF data sources to be more scalable when adding other kinds of peers and services to the network. Edutella use super-peer based topology, in which peers are organized in 'HyperCup' [33]. to route queries. This project planned to use the same topology. However, we intended to construct multiple data views based on superpeers' index. The index contains information about metadata usage at each peer. While the views can be use to identify the right data source when information form several data sources need to be combine during data retrieval.

4.4. PIAZZA

The Piazza Peer Data Management Project (Piazza PDMS) [35] address the issue of heterogeneity in P2P systems on schema level, which allows for information sharing with different schemas relying on local mappings between schemas. The vision of Piazza is to blend the extensibility of Hypertext Markup Language (HTML) with the semantics of data management applications. Peer-Programming Language (PPL) [17] is used for mediating between peer schemas, which uses Global-Local-as-View (GLAV) [28] formalism to specify local mappings. GLAV is one of the data source modeling approach, i.e., how their content is related to the global schema mediator. GLAV is the combination of Global-as-View (GAV) [25] [38] and Local-as-View (LAV) [25]. By using GAV, the mediated schema is defined as a set of views over the data sources. GAV specifies how to extract tuples for the mediated schema from the source. While in LAV, the content of data sources

are describe as views over the mediated schema. With these schemas mapping concepts, extending the schema with a new source is problematic because the new source may have impact on the view definition of the mediated schema construct. Whereas, this view definition is essential to tell how to use the sources in order to retrieve data. Ideally, P2P systems must adapt the scenario of extending the schema from new sources due to the freedom for peers to join-in and withdrawal from the P2P community. In contrast to this project, mapping is done based on shared knowledge-based (such as ontology [14], semantic dictionary [2], concept hierarchy [15], mapping table [4]), instead of global or local schema of participated peers. The use of shared knowledge will enable all participated peers to communicate about domain of discourse without necessarily operating on a globally shared mediation[14].

5. PROPOSED FRAMEWORK REQUIREMENTS

In order to generate a semantic-based personalize multiple views in P2P database integration, several main requirements must be met.

5.1. INFORMATION SEARCHING

It is a natural aspiration in the P2P network to imagine that the query is not blindly broadcasted. Thus, P2P network must include efficiency in information searching. Some improvements have been done in analyzing the right peers to send the query [16] [9], where query is routed intelligently. During the routing process, new logical links will be added in a dynamic way when related peers join. At the same time, dead logical link are removed when peers leave. Theoretically, this allows large networks to be built rapidly. However, most of the experiments are based on static semantic topologies [16] and small-scale environment [9]. It is essential to do some experiments in dynamic semantic topologies and larger scale environment.

5.2. MAPPING COMPUTATION

Addressing the issue on how to compute mapping among database schemas, firstly we must understand how to compute a mapping that represent a path in order to enable reformulating a query posed on one peer to queries on its neighbours (i.e., composition mappings [28] and revisable schema transformation [30]). Then, we need to show how to compute mapping that considering the semantic aspects as in [41] [24]. Kementsietsidis.et.al has stated that *mapping table* is appropriate for data mapping in a P2P environment [24]. However their experiment has been design to work with peers with

local relational model (LRM) as a data model.

Bernstein.et.al claimed that LRM is design specifically for P2P application [4]. The fact that data model in P2P is not limited to relational model, which is structured data model must be accepted. As a result, there might be any other approach for data mapping need to be explored especially for unstructured data.

5.3. KNOWLEDGE EXTRACTION

Due to the dynamic nature of P2P, it is impractical to have a global knowledge among peers. Consequently, we should fully utilize peers' local knowledge of peers. So the peers' local knowledge must be adaptive to the dynamic environment. Due to the fact that available information is often heterogeneous and distributed, an efficient peers' knowledge extraction is required in order to fulfil demanding for complete access to available information. Once the information source has been found, access to data in the source has to be provided. This means that each information sources have to work together with the system that is querying the information. In order to achieve interoperability among peers, the *meaning* of the information interchanged has to be understood across the peers' community. Uschold and Gruninger stated that interoperability as a key application of ontologies in order to achieve interoperability among different modelling methods, paradigms, languages and software tools in information integration [39]. The role of ontology can be roughly categorized as [40]: 1) Explicit description of information source semantics. 2) Global query model and 3) Verification of integration description. Therefore, ontology is needed to solve the problem of interoperability between data source schemas across different peers, providing a shared understanding of common domains.

Today, various needed ontologies are scattered over various peers. Thus, ontologies, if shared among the interoperating peers, allow the exchange of data to take place not only at a syntactic level, but also at a semantic level. This is due the fact that no single and universal ontology can be built has been agreed by Semantic Web community [14]. Therefore, a key challenge in knowledge extraction is not only enabling the interoperability between data sources but also among different ontologies.

5.4. KNOWLEDGE BASED FOR GENERATING VIEWS

Some knowledge is required for generating views. This knowledge should be stored in some peers' repository. The stored knowledge is obtained by analyzing the meta-data of the data source from

semantic mapping that has been computed from respective peers. In federated systems, the knowledge for generating view is built in bottom-up approach where the meta-data information is extracting from shared databases and merging the meta-data incrementally [23]. Since the data source in P2P is volatile, multiple 'dynamic knowledge generator' is essential for producing multiple different views.

6. A PEER ARCHITECTURE

Figure 4 prescribe the preliminary design of application framework for peer architecture. The components are local data source, knowledge source, meta-data extractor and integrator, local peer repository, user interface, query manager and advertiser.

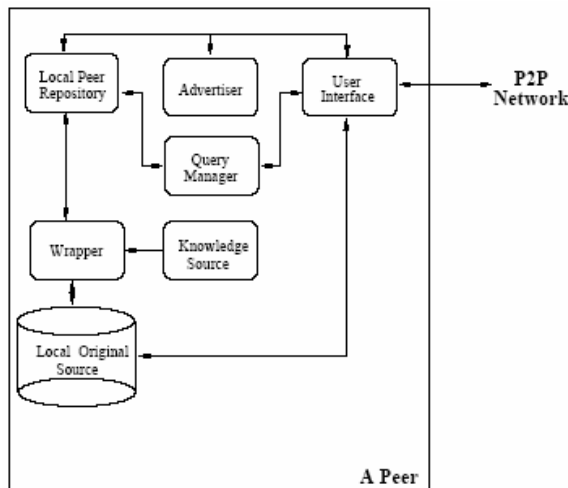


Fig. 4 – A Peer Architecture of the Application Framework

6.1. LOCAL DATA SOURCE

Peers may have local databases as local sources of information. This is where a peer can physically store information to be shared with other peers on network.

6.2. KNOWLEDGE SOURCE

The purpose of knowledge source is to provide some intelligence during extraction of Local Original Source and transforming the result to Local Peer Repository. The schema that is being extracted need to be analyzed, clustered then merged. Schema analyzing is needed to determine implicit relationship between schemas. Clustering is should be done to classify the schema in order to determine their similarities. Merging process is used to merge the schema structure with respect to the original schema terms. These three processes require some knowledge for constructing the best integrated data

in Local Peer Repository. This knowledge may be in the form of ontology [14], semantic dictionary [2], concept hierarchy [23], mapping table [4] or else.

6.3. WRAPPER

This component responsible for the extraction of internal meta-data sources and transform into the data format used in Local Peer Repository. This task comprises (i) access Local Data Source, (ii) start resource discovery process, (iii) filter-out some data that is duplicate or incomplete, (iv) integrate / merge similar data sources, (v) transform the integrated data into a specified data model (such as RDF-S, XML schema) into Local Peer Repository. This project intended to use knowledge-based mechanism such as ontologies, concept hierarchies and semantic dictionaries as a knowledge source in order to capture semantics of schema during transformation process.

6.4. Local Peer Repository

This component stores the integrated schema in local peers' format / data model. It contents can be used for query processing. In order to manage its data model and views as well as data acquired from other peers on network, each peer must have Local Peer Repository that must be able to (i) mediate between views and stored information and (ii) support query formulation and processing.

6.5. USER INTERFACE

The User Interface of the peer provides individual views on the information available in local sources. Besides that, this component is used to allow user to edit their Local Original Source and Local Peer Repository, sending advertisement about the data they can provide, as well as to easily formulate queries.

6.6. QUERY MANAGER

Query Manager is the coordinating component controlling the process of distributing queries. It receives queries from the User Interface or forwarded queries from other peers. Either it tries to answer the query or distribute it further according to the content of the query.

6.7. ADVERTISER

The task of Advertiser is to proactively advertise the available data of a peer in the P2P network. Since this project is based on JXTA platform, we planned to enhance the capabilities of existing Advertisement component in JXTA.

7. DISCUSSION AND CONCLUSION

P2P itself is an immature technology with many unanswered questions. With the idea of sharing distributed, heterogeneous resources over a network, P2P is a revival paradigm. Even if the WWW contains more information than any single P2P community, it cannot substitute that P2P data management system (PDMS) because it provides a potential architecture to support knowledge-based searching, where semantic of information is considered. Adopting multiple views on PDMS is an important trend that will combine the P2P and semantic web technologies together to achieve both better precision in information search and fulfil the different users' need at the same time.

REFERENCES

- [1] M. B. Al-Mourad, W. A. Gray, and N. J. Fiddian. Multiple views with multiple behaviours for interoperable object-oriented database systems. Proceeding of 14th International Conference on Database and Expert Systems Applications Czech Republic, 2003. Lecture Notes in Computer Science (LNCS2736), Springer-Verlag, 2003.
- [2] M. B. Al-Mourad, W. A. Gray, and N. J. Fiddian. Semantically rich materialisation rules for integrating heterogeneous databases. 22nd British National Conference on Databases, BNCOD 22, Sunderland, England, UK, July 2005.
- [3] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, May 2001.
- [4] P. A. Bernstein, F. Giunchiglia, A. Kementsietsidis, J. Mylopoulos, L. Serafini, and I. Zaihrayeu. Data management for peer-to-peer computing: A vision. Proceedings of the Fifth International Workshop on the Web and Databases (WebDB 2002), 2002.
- [5] I. Brunkhorst, H. Dhraief, A. Kemper, W. Nejdl, and C. Wiesner. Distributed queries and query optimization in schema-based p2p-systems. The VLDB Journal The International Journal on Very Large Data Bases, September 2003.
- [6] D. Calvanese, G. D. Giacomo, M. Lenzerini, and R. Rosati. Logical foundations of peer-to-peer data integration. Proceedings of the twenty-third ACM SIGMOD-SIGACTSIGART symposium on Principles of database systems PODS '04, 2004.
- [7] J. C. P. Carvalho and A. S. da Silva. Finding similar identities among objects from multiple web sources. Proceedings of the 5th ACM international workshop on Web information and data management (WIDM'03, pages 90–93, November 2003.
- [8] M. A. Casanova and V. M. P. Vidal. Towards a sound view integration. Proceedings of the 2nd ACM SIGACT-SIGMOD symposium on Principles of database systems, pages 36–47, 1983.
- [9] H. Ding, I. T. Solvberg, and Y. Lin. A vision on semantic retrieval in p2p network. Proceedings of the 18th International Conference on Advanced Information Networking and Applications (AINA 2004), March 2004.
- [10] A. Doan, P. Domingos, and A. Levy. Learning source descriptions for data integration. Proceedings of 3rd international Workshop on the Web and Databases, 2000.
- [11] M. Ehrig, C. Tempich, J. Broekstra, F. van Harmelen, M. Sabou, R. Siebes, S. Staab, and H. Stuckenschmidt. Swap: Ontology-based knowledge management with peer-to-peer technology. Proceedings of the 1st National Workshop Ontologie-basieres Wissensmanagement (WOW2003), 2003.
- [12] L. Gong. Jxta: A network programming environment. *IEEE Internet Computing*, 5:88–95, 2001.
- [13] L. Gong. Project jxta: A technology overview. Technical report, Sun Microsystems, Inc., 901 San Antonio Road, Palo Alto, CA 94303 USA, October, 29 2002.
- [14] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. International Workshop on Formal Ontology, Padova, Italy. Revised August 1993. Published in *International Journal of Human-Computer Studies*, Volume 43, Issue 5-6 Nov./Dec. 1995, Pages: 907-928, special issue on the role of formal ontology in the information technology, 1993.
- [15] P. Haase, J. Broekstra, M. Ehrig, M. Menken, P. Mika,
- [16] M. Plechawski, P. Pyszlak, B. Schnizler, R. Siebes, S. Staab, and C. Tempich. Bibster – a semantics-based bibliographic peer-to-peer system. Proceedings of the International Semantic Web Conference (ISWC2004), Hiroshima, Japan, November 2004.
- [17] P. Haase and R. Siebes. Peer selection in peertopeer networks with semantic topologies. In *International Conference on Semantics of a Networked World: Semantics for Grid Databases*, Paris, 2004.
- [18] A. Y. Halevy, Z. G. Ives, D. Siciu, and I. Tatarinov. Schema mediation in peer data management systems. *International Conference on Data Engineering*, March 2003.
- [19] R. Hull. Managing semantic heterogeneity in databases: a theoretical prospective.

- Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principle of database systems, 1997.
- [20] R. Hull and G. Zhou. A framework for supporting data integration using the materialized and virtual approaches. ACM SIGMOD Record, Proceedings of the 1996 ACM SIGMOD international conference on Management of data SIGMOD '96, 25(2):481–492, 1996.
- [21] E. Jeong and C. N. Hsu. Induction of integrated view for xml data with heterogeneous dtDs. Proceedings of the tenth international conference on Information and knowledge management, pages 151–158, October 2001.
- [22] M. Karnstedt, K. Hose, and K. U. Sattler. Query routing and processing in schema-based p2p systems. Prooceding of 15th International Workshop on Database and Expert Systems Applications, pages 544–548, 30 August – 3 September 2004.
- [23] D. D. Karunaratna. Exploration of Semantic Information in the Creation of Multiple views of a Federated Database System. PhD thesis, Department of Computer Science, Cardiff University, May 2000.
- [24] D. D. Karunaratna, W. A. Gray, and N. J. Fiddian. Organizing knowledge of a federation database system to support multiple view generation. Proceedings of the 15th KRDB Workshop Seattle, WA, 10, May 31 1998.
- [25] A. Kementsietsidis and R. J. M. M. Arenas. Mapping data in peer-to-peer systems: Semantics and algorithmic issues. Proceedings of the 2003 ACM SIGMOD international conference on Management of data, 2003.
- [26] I. Koffina, G. Serfiotis, and V. Christophides. Foundation for information integration: A state of the art. Technical report, DELOS: Network of Excellence on Digital Libraris. Information Society Technologies, 2005.
- [27] S. Lawrence. Net news: Search engines fall short. Science Online, 285(5426):295, 16 Jul 1999.
- [28] W. Litwin, L. Mark, and N. Roussopoulos. Interoperability of multiple autonomous databases. ACM Computing Surveys, 22(3):267–291, 1990.
- [29] J. Madhavan and A. Y. Halevy. Composing mappings among data sources. In proceedings of VLDB, 2003.
- [30] P. Martin and W. Powley. Multidatabase systems: Database integration using multidatabase views. Proceedings of the 1993 conference of the Centre for Advanced Studies on Collaborative research: distributed computing, 2, October 1993.
- [31] P. McBrien and A. Poulouvasilis. Data integration by bidirectional schema transformation rules. In Proceedings of ICDE03, pages 227–238, 2003.
- [32] W. Nedjl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmer, and T. Risch. Edutella: A p2p networking infrastructure based on rdf. Proceedings of the eleventh international conference on World Wide Web, May 2002.
- [33] W. Nejdl, W. Siberski, and M. Sintek. Design issues and challenges for rdf and schema-based peer-to-peer systems. ACM SIGMOD Record, 32, September 2003.
- [34] M. Schlosser, M. Sintek, S. Decker, and W. Nejdl. Hypercup – hypercubes, ontologies and efficient search on p2p networks. 1st Workshop on Agents and P2P Computing, Bologna, Italy, July 2002.
- [35] D. Sullivan. Search engine sizes. Technical report, Search Engine Watch, Jan 28 2005.
- [36] I. Tatarinov, Z. Ives, J. Madhavan, A. Halevy, D. Suci, N. Dalvi, X. Dong, Y. Kadiyska, G. Miklau, and P. Mork. The piazza peer data management project. SIGMOD Record, 32(3), September 2003.
- [37] S. Tejada. Learning Object Identification Rules for Information Integration. PhD thesis, Faculty of the Graduate School, University of Southern California, August 2002.
- [38] S. Tejada, C. A. Knoblock, and S. Minton. Learning object identification rules for information integration. Information Systems, 26(8):607–635, December 2001.
- [39] J. Ullman. Information integration using logical views. Theoretical Computer Science, 239(2):189–210, 2000.
- [40] M. Uschold and M. Gruninger. Ontologies: Principles methods and applications. Knowledge Engineering Review, 11(2):93–155, 1996.
- [41] H. Wache, T. Voegelé, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Hubner. Ontology-based integration of information – a survey of existing approaches. In Proceedings of IJCAI-01 Workshop: Ontologies and Information Sharing, pages 108–117, 2001.
- [42] C. Yu and L. Popa. Semantic adaptation of schema mappings when schemas evolve. Proceedings of the 31st international conference on Very large data bases VLDB '05, August 2005.



Mohamed Basel Al-Mourad received his BSc in Informatics Engineering, from Aleppo University, Syria, and his PhD degree from Cardiff University of Wales, United Kingdom. Currently Dr. Al-Mourad is a Senior Lecturer of Information Systems at College of Information Technology, University of Dubai Dubai, UAE.

Rozalina Mohammed received her BSc degree and her Masters from the Computer Science Department, University Technology Malaysia in 1996 and 1998 respectively. Ms. Mohammed is currently working on her PhD at Aston University, Birmingham, UK



Yaser M.A. Khalifa received his BSc degree from Alexandria University, Egypt, and his PhD degree from Cardiff University of Wales, United Kingdom, in 1992 and 1997 respectively. Dr Khalifa is currently the Vice President, Director of Technology, at General Energy Services, Kent Lakes, New York, USA. Dr. Khalifa's primary research interest is in the application of evolutionary algorithms in the design of analog and digital circuits, and smart power management systems.