A Vision to Construct Multiple Data Views in Peer Data Management Systems

Rozlina Mohamed, M. Basel Al-Mourad and [1]Yasser M. A. Khalifa

Aston University, Computer Science Dept. B4 7ET UK
[1] Electrical and Computer Eng., State University of New York NY 1261-2443
(mohamedr, m.b.al-mourad)@aston.ac.uk
[1] yaserma@engr.newpaltz.edu

Peer data management systems (PDMS) has made available to almost all computer users to have a basic physical capability of sharing data. Meanwhile, problem of data sharing from multiple locations has raised significant attention since the *actual* data is stored in users' data sources at peers' location [KP05]. Usually, a logical virtual mediated schema is used to create *data views* that describe contents of the data sources. View is a stored query, which is formulated over data sources schemas in order to get access to the actual data sources. Thus, any user query is formulated over this view.

There are several well-known approaches to create view for data integration, i.e Global-As-View (GAV), Local-As-View (LAV), Global-Local-As-View (GLAV) and Both-As-View (BAV) [MP03] [MH03]. These approaches are originally developed for federated database management systems (FDBMS) or mediated systems, it require for extension in order to apply in p2p environment. Meanwhile, Local Relational Model (LRM) [BGK+02] is a data model specifically design for p2p application. LRM directly combined related peers to each other via domain relation. However, it is limited to relational databases shared among peers. On the other hand, LRM is proposed on pure p2p network. In contrast, this research intended to be applied on super-peer network in order to provide better scalability than pure p2p network. For now, PDMS such as SWAP [ETB+03], Bibster [HBM+04], Edutella [NWQ+02], Piazza [TIM+03] and AutoMed [BKL+04]. Piazza and AutoMed have highlight the use of view in PDMS. The GLAV and composition mapping is initiate in Piazza project to optimize query processing by pre-compose the common cases of mapping. However, they are lacking on deciding which path to pre-compose and ensure the optimizer uses the composition appropriately in order to provide a good optimization plan that essential to optimize the query processing.

Then, AutoMed has brought-in the BAV approach to construct single shared view among peers in super-peer node. However, having single views is not flexible to support difference peers' need. Different peers have different reason for sharing data, and the same peer might

want to integrate the same distributed information in various ways to suit different roles and tasks. Therefore, multiple data views is required to full different peers' need. Numbers of significant reasons identified for constructing multiple data views in p2p environment are

(1) Applicable approach to identify relevant sources of required information in a very limited source of knowledge

(2) Convey semantic meaning of information. View has been constructed based on users' need. This will lead to the integration of required information only and

(3) Reduce burden of query processing. View may direct for local query processing, where source of each required information is identified before sub-query being posted to other peers.

In order to full the ambitious aim, several requirements has been identified as follows:

(1) information searching technique to avoid network flooding

(2) semantic mapping approach to map the peers' advertisement to index and query to local peers' schema

(3) extracting knowledge strategy to identify location of required data and

(4) knowledge-based mechanism to handle semantic mapping and constructing multiple different data views [MAMY06].

Research on multiple data views that has been done concerned on materialize views developed in FDBMS [AMGF05]. However, the FDBMS environment is dissimilar to p2p settings, where view in p2p should be constructed in the absence of central or global knowledge.

Bibliography

[AMGF05] M. B. Al-Mourad, W. A Gray, and N. J. Fiddian. Semantically rich materialisation rules for integrating heterogeneous databases. 22nd British National Conf. on DBs,BNCOD 22, 2005.

[BGK+02] P. A. Bernstein, F. Giunchiglia, A. Kementsietsidis, J. Mylopoulos, L. Sera̅ni, and I. Zaihrayeu. Data management for p2p computing: A vision. Proc. of the 5th Int. Workshop on the Web and DBs, WebDB, 2002.

[BKL+04] M. Boyd, S. Kittivoravitkul, C. Lazanitis, P.J. McBrien, and N. Rizopoulos. Automed: A bav data integration system for heterogeneous data sources. Proc. of CAiSE04,Springer Verlag LNCS, 3084:82{97, 2004.

[ETB+03]   M. Ehrig, C. Tempich, J. Broekstra, F. V. Harmelen, M. Sabou, R. Siebes, S. Staab, and H. Stuckenschmidt. Swap: Ontology-based knowledge management with p2p technology. Proc. of the 1st National Workshop Ontologie-basiertes Wissensmanagement (WOW2003), 2003.

[HBM+04]   P. Haase, J. Broekstra, M.Ehrig, M. Menken, P.Mika, M. Plechawski, P. Pyszlak, B. Schnizler, R. Siebes, S. Staab, and C. Tempich. Bibster - a semantics-based bibliographic p2p system. Proc. of the Int. Semantic Web Conference (ISWC2004), 2004.

[KP05]   G. Koloniari and E. Pitoura. P2p management of xml data: Issues and research challenges. ACM Sigmod, Vol. 34, No. 2, 34(2), June 2005.

[MAMY06]R. Mohamed, M. B. Al-Mourad, and Y. M. A. Yasser. Towards multiple views in peer data management systems. Proc. of ICCA 2006, 4th Int. Conf. on Computer Application, February 23-24 2006.

[MH03]   J. Madhavan and A. Y. Halevy. Composing mappings among data sources. Proc. of VLDB, 2003.

[MP03]   P. McBrien and A. Poulovassilis. De¯ning p2p data integration using both as view rules. Proc. Workshop on DBs, IS and P2P Computing, at VLDB'03, Berlin, September 2003.

[NWQ+02] 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. Proc. of the 11th Int. Conference on World Wide Web, May 2002.

[TIM+03]   I. Tatarinov, Z. Ives, J. Madhavan, A. Halevy, D. Suciu, N. Dalvi, X. Dong, Y. Kadiyska, G. Miklau, and P. Mork. The piazza peer data management project. SIGMOD Record, 32(3), September 2003.