Peering and Querying e-Catalog Communities

Boualem Benatallah[1], Mohand-Said Hacid[2], Hye-young Paik[1]
Christophe Rey[3] and Farouk Toumani[3]

[1] CSE, University of New South Wales, Australia, {boualem,hpaik}@cse.unsw.edu.au
[2] LIRIS, University Lyon I, France, mshacid@liris.univ-lyon1.fr
[3] LIMOS, ISIMA, University Blaise Pascal, France, {rey,ftoumani}@isima.fr

More and more suppliers are offering access to their *product or information portals* (also called e-catalogs) via the Web. The key issue is *how to efficiently integrate and query large, intricate, heterogeneous information sources such as e-catalogs*. Clearly, traditional data integration approach, where the development of an integrated schema requires the understanding of both structure and semantics of all schemas of sources to be integrated, is hardly applicable because of the dynamic nature and size of the Web. We present *WS-CatalogNet*: a Web services based data sharing middleware infrastructure whose aims is to enhance the potential of e-catalogs by focusing on scalability and flexible aspects of their sharing and access. The salient features of WS-CatalogNet are:

(i) Ontological Organization: We use the concept of *e-catalog community* as a way of organizing and integrating a potentially large number of dynamic e-catalogs [3]. An e-catalog community is a container of catalogs that offer products of a common domain (e.g., community of laptops). It provides an ontological description of desired products without referring to any actual provider (e.g., Dell.com). Communities of e-catalogs are established through the sharing of high-level meta-information. Actual providers can register with any community of interest to offer the desired products. E-catalog providers can join or leave any community of interest at any time. Communities meaningfully organize and divide the information space into groups of manageable spaces (e.g., putting similar products together).

(ii) Decentralized Information Sharing: Existing e-catalog organization techniques use centralized categorization and indexing schemes, whereas the participating e-catalogs are distributed and autonomous. Given the highly dynamic and distributed nature of e-catalogs, novel techniques involving peer-to-peer centric categorization and indexing schemes will become increasingly attractive. We use *peer relationships* among e-catalog communities to allow decentralized sharing of catalog information. Query routing between communities is based on such relationships (with no or minimal knowledge of the schema about the other party). The objective is to achieve scalable information sharing and access through the incremental meta-data driven discovery and formation of interrelationships between e-catalog communities.

(iii) Flexible Selection of Relevant e-Catalogs: Because of the variety of e-catalogs offering similar information and the large number of available e-catalogs, it is important to provide appropriate support to first select those e-catalogs that are relevant to a given query. Also, catalog selection techniques should support flexible matching since it is unrealistic to expect queries and catalog descriptions to exactly match. In *WS-CatalogNet*, a query is posed over a community ontology. We formalize relevant e-catalog selection as a new instance of the query rewriting problem [1], where a query Q over a community C is reformulated in terms of (i) Q_{local}: local queries (i.e., the part of Q that can be answered by e-catalogs registered with C), (ii) Q_{rest}: remaining parts of Q that cannot be answered by C. Q_{rest} part will be forwarded to the other communities known to C via the peer relationships. The expected answer from forwarding is the identification of external (i.e., registered to other community) e-catalogs who can resolve the query. Once the relevant e-catalogs are identified, the community sends the query to the e-catalogs and assemble the results. We provide a formalization of query rewriting in the context of category hierarchy based ontologies and propose a hypergraph-based algorithm to effectively compute best rewritings of a given request.

References

[1] F. Baader, R. Küsters, and R. Molitor. Rewriting Concepts Using Terminologies. In *KR'00, Colorado, USA*, pages 297–308, April 2000.

[2] A. Bouguettaya, B. Benatallah, L. Hendra, M. Ouzzani, and J. Beard. Supporting Dynamic Interactions among Web-based Information Sources. *IEEE TKDE*, 12(5):779–801, Sept/Oct 2000.

[3] H. Paik, B. Benatallah, and R. Hamadi. Dynamic Restructuring of E-Catalog Communities Based on User Interaction Patterns. *WWW Journal*, 5(4):325–366, 2002.