OBJECTIVES

The DataRing project addresses the problem of P2P data sharing for online communities, by offering a high-level network ring across distributed data source owners. Users may be in high numbers and interested in different kinds of collaboration and sharing their knowledge, ideas, experiences, etc. Data sources can be in high numbers, fairly autonomous, i.e. locally owned and controlled, and highly heterogeneous with different semantics and structures.

METHODOLOGY AND RESULTS

Our approach is to organize community members in a P2P architecture where each member can share data with the others through a P2P overlay network. The project built two integrated demos to exercise different DataRing services that emphasize P2P data sharing and collaboration.

A Social-based P2P Data Sharing System (P2PShare)

P2PShare is a P2P system for large-scale probabilistic data sharing, particularly in scientific communities. It takes into account heterogeneous data, and leverages content-based and expert-based recommendation for discovering the data relevant to queries.

Main components

ProbDB (http://probdb.gforge.inria.fr) is a probabilistic database system built on top of a classical DBMS. Instead of directly modifying the DBMS and adding "native" primitives to it, we have chosen to implement ProbDB on top of the DBMS, and thus be able to change the underlying DBMS with a slight programming effort.

WebSmatch (http://websmatch.gforge.inria.fr) is a flexible environment for Web data integration, based on real, end-to-end data integration scenarios (e.g. over public data or scientific data). WebSmatch supports the full process of importing, refining and integrating data sources and uses third party tools for high quality visualization.

P2Prec (http://p2prec.gforge.inria.fr) is a social-based P2P recommendation system for large-scale content sharing that leverages content-based and social-based recommendation. The main idea is to recommend high quality documents related to query topics, by exploiting friendship networks.

SON (http://www-sop.inria.fr/teams/zenith/SON) is an open source development platform for P2P networks based on P2P and SOA concepts. SON components communicate by asynchronous message passing to provide weak coupling between system entities. To scale up and ease deployment, we rely on a Distributed Hash Table (DHT) for publishing and discovering services or data.

P2PShare demo

CONCLUSION

The project produced visible software prototypes that implement high-level services (query, trust and privacy, semantic data integration, recommendation, etc.). All critical aspects of the project (data model, query processing, data integration, data privacy and trust, recommendation) led to high-quality publications in very good journals (PVLDB, IEEE TKDE, IS) and very good conferences (SIGMOD, CIKM, EDBT, ICDT, IEEE P2P, CAiSE, etc.).

The research done in DataRing has been carried on in other projects of the participants, in particular, the WebDam ERC and the BigDatanet project of the Zenith team.