RACCOON: A Peer-Based System for Data Integration and Sharing

Chen Li, Jia Li, and Qi Zhong
Information and Computer Science, University of California, Irvine, CA 92697, USA
{chenli,jiali,qzhong}@ics.uci.edu

1. Introduction

Recent database applications see the emerging need to support data integration in distributed, peer-to-peer environments [1], in which autonomous peers (sources) connected by a network are willing to exchange data and services with each other. To address related research challenges, we are developing a system called "RACCOON," which allows different sources to integrate and share their data.

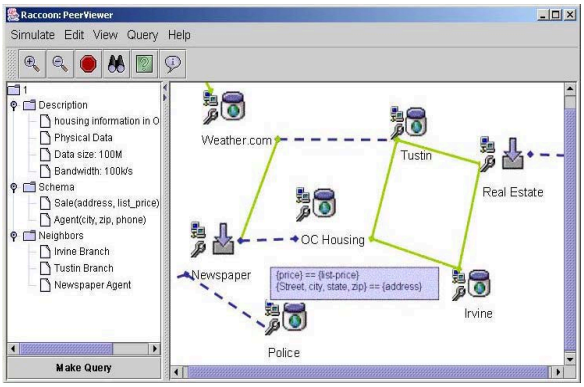

Figure 1. PeerViewer.

Compared to existing peer-based DBMS systems [2, 3, 4, 5], the RACCOON system has the following features. (1) It uses a rich language to describe various source contents. The language annotates each source with certain attributes such as its data size, network bandwidth, querying interfaces, and whether it is "active" or "passive." It also describes source schema information and possible constraints. (2) The system uses two kinds of links in the peer network. The first kind, called "propagation links," is used to support searches for sources similar to a peer relation. The second, called "mapping links," is used to define semantic mappings between two sources. (3) During a search, the system uses schema-mapping techniques to find relations similar to the given relation and automatically suggests a semantic mapping between the two relations. Thus it can reduce users' efforts to create mappings. A mapping between a peer A and a peer B is defined in the format: $Q_A\ op_m\ Q_B$, where Q_A and Q_B are queries on databases of A and B, respectively. The mapping operator op_m describes the relationship between the query results, such as inclusion ($\subseteq$) or equality ($=$). (4) To meet different user needs to search for information, the system supports two different querying modes: focused querying mode (querying on specific sources) and extended querying mode (the system will use semantic mappings to expand the query to other sources). (5) It provides a visualization toolkit called "PeerViewer" to allow users to browse the contents of peers in the network, create links, and query sources.

1.1. Demonstration

We will use an application to show several important features of the RACCOON system. We will demonstrate the visualization tool PeerViewer, a screenshot of which is shown in Figure 1. The user can explore the network with the tool. We will show how the system finds peer relations similar to a given relation using schema-mapping techniques. The system will also suggest semantic mappings for the user to choose. We will show the two different querying modes, particularly how a query is expanded using the semantic mappings in the extended querying mode to compute as many answers to the query as possible.

This work was supported by NSF CAREER award No. IIS-0238586. The project Web page is at `http://www.ics.uci.edu/~raccoon/`.

References

[1] P. A. Bernstein et al. Data management for peer-to-peer computing: A vision. In *WebDB*, 2002.
[2] A. Y. Halevy et al. Schema mediation in peer data management systems. In *ICDE*, 2003.
[3] T. Katchaounov et al. Scalable view expansion in a peer mediator system. In *DASFAA*, 2003.
[4] W. S. Ng et al. PeerDB: A P2P-based system for distributed data sharing. In *ICDE*, 2003.
[5] The Hyperion Project. http://www.cs.toronto.edu/db/hyperion/.