OBIRE: Ontology Based Bibliographic Information Retrieval in P2P Networks

Xiangyu Liu, Brunel University, UK
Maozhen Li, Brunel University, UK
Yang Liu, Brunel University, UK
Man Qi, Canterbury Christ Church University, UK

ABSTRACT

It has been widely recognized that bibliographic information plays an increasingly important role for scientific research. Peer-to-peer (P2P) networks provide an effective environment for people belonging to a community to share various resources on the Internet. This paper presents OBIRE, an ontology based P2P network for bibliographic information retrieval. For a user query, OBIRE computes the degree of matches to indicate the similarity of a published record to the query. When searching for information, users can incorporate their domain knowledge into their queries which guides OBIRE to discover the bibliographic records that are of most interest of users. In addition, fuzzy logic based user recommendations are used to compute the trustiness of a set of keywords used by a bibliographic record which assists users in selecting bibliographic records. OBIRE is evaluated from the aspects of precision and recall, and experimental results show the effectiveness of OBIRE in bibliographic information retrieval.

Keywords: Fuzzy Logic, Information Retrieval, Ontology Merging, P2P, Semantic Web

INTRODUCTION

The past few years have witnessed a rapid development of peer-to-peer (P2P) networks for users belonging to a community to share various resources over the Internet (Milojicic et al., 2002). P2P networks usually organize peer nodes in a decentralized way to enhance reliability in resource sharing. Resources can be arbitrarily distributed into peer nodes without a structure, or they can be distributed following certain structures such as Distributed Hash Tables (DHTs). DHT based P2P networks such as Chord (Stoica et al., 2002), Pastry (Rowstron & Druschel, 2001), CAN (Ratnasamy et al., 2001) have shown enhanced scalability in routing lookup messages for locating resources with a guaranteed number of hops.

It has been widely recognized by the research community that bibliographic information plays an increasingly important role for scientific research. The salient features of
P2P networks make them a good candidate for scalable and reliable bibliographic information sharing wherein peer researchers can effectively exchange bibliographic information to promote research. However, two challenging issues have to be dealt with when applying P2P technologies for bibliographic information retrieval – the support of range queries and the integration of peer expertise knowledge in queries.

Although DHT structured P2P networks can achieve high scalability in routing lookup messages, they only support exact matches of resources using single hash keys. Based on DHTs, some works are available that support range queries in P2P networks. For example, SWORD (Oppenheimer et al., 2004) supports multi-attribute range queries to locate suitable peer nodes. In SWORD, each DHT is used to manage a resource attribute. A query is routed by identifying its preferred attribute and using the corresponding DHT. Similarly, Building on Chord, MAAN (Cai et al., 2004) also uses multiple DHTs to manage range queries. Mercury (Bharambe, Agrawal, & Seshan, 2004) is another example that supports range queries over multi-attributes. Like SWORD and MAAN, Mercury maintains a separate logical overlay for each attribute albeit it does not employ DHTs. Maintaining a DHT or logical overlay for each resource attribute involves updating each of them when a resource advertisement is received. As the number of attributes of resources increases, the update traffic would also increase heavily.

On the other hand, when searching for bibliographic information in a P2P network, users would expect to receive a list of bibliographic records that maximally satisfies their requirements. To meet this end, the expert knowledge of users in a particular research domain needs to be reflected in a P2P network guiding the system to retrieval the information that is most relevant to a user query. With the development of Semantic Web (Berners-Lee, Hendler, & Lasilla, 2001), ontologies have been widely used for knowledge representation. Knowledge related to a research domain can be represented in the form of ontologies which can be considered as global ontologies. The expert knowledge of users can be considered as user ontologies. It is still a challenging task for P2P networks to automatically merge global ontologies with user ontologies so that users can be maximally satisfied when searching for bibliographic information (Kalfoglou & Schorlemmer, 2003).

This paper presents OBIRE, an ontology based P2P network for bibliographic information retrieval. A novelty of OBIRE lies in its capability to merge global ontologies with user ontologies to maximally satisfy users in their queries. OBIRE supports three types of matches – keyword exact matches, matches with semantics, and matches with merged ontologies. In addition, fuzzy logic based user recommendations are used to compute the trustiness of a set of keywords used by a publication which assists users in selecting bibliographic records. Experimental results show the performance of OBIRE in bibliographic information retrieval from the aspects of the three types of matches it supports.

The remainder of the paper is organized as follows: we present OBIRE with a focus on matchmaking of bibliographic records, ontology merging, and fuzzy logic based user recommendations. Next, the paper evaluates the performance of OBIRE. After that, we discuss some related work from the aspects of bibliographic information retrieval and ontology merging. Finally, we conclude the paper and point out some future work.

**OBIRE**

OBIRE adopts a super-peer structure to manage peers that can be further grouped based on their research interests. Figure 1 shows the components of OBIRE. Peers publish bibliographic records to a bibliographic information base (BIB). Each super-peer nodes manages one BIB and super-peers exchanges bibliographic records periodically. For a research domain, OBIRE maintains an ontology representing global domain knowledge. When searching for bibliographic information, peers can incorporate
Related Content

**A Distributed Storage System for Archiving Broadcast Media Content**
[www.igi-global.com/chapter/distributed-storage-system-archiving-broadcast/20516?camid=4v1a](www.igi-global.com/chapter/distributed-storage-system-archiving-broadcast/20516?camid=4v1a)

**Scalable Distributed Two-Layer Data Structures (SD2DS)**
[www.igi-global.com/article/scalable-distributed-two-layer-data/78151?camid=4v1a](www.igi-global.com/article/scalable-distributed-two-layer-data/78151?camid=4v1a)
Functional Genomics Applications in GRID
Luciano Milanesi, Ivan Merelli, Gabriele Trombetti, Paolo Cozzi and Alessandro Orro
(pp. 899-917).
www.igi-global.com/chapter functional-genomics-applications-grid/64521?camid=4v1a

Service and Management Oriented Traffic Information Grid
www.igi-global.com/chapter/service-management-oriented-traffic-information/64454?camid=4v1a