Semantic Search in Peer-to-Peer based Digital Libraries

Hao Ding
Dept. of Computer and Information System, Norwegian Univ. of Sci. &Tech.
Sem Sælands vei 7-9, NO-7491, Trondheim, Norway
hao.ding@idi.ntnu.no

ABSTRACT

This paper gives an outline of a PhD thesis inspired by the information searching over heterogeneous metadata records within the scenario of Peer-to-Peer (P2P) based digital libraries. We survey a representative set of P2P systems in order to build prospective platform and discuss why a super-peer (peer community) based topology is suitable for schema-based P2P applications. Furthermore, we exploit ontologies for storing metadata records in a structured and formal way, and for broadening queries with closely related terms, thus yielding more relevant results. All metadata records into RDF format and a RDF database, such as Sesame, is adopted to store and query these records and schema information. Additionally, since words generally have more than one sense, we have to choose the right sense and highly related terms for extending the query. Such information can be delivered by applying a lexical database, like WordNet, which models terms with their meanings and relationships in between. The resulting prototype system is expected to serve as a framework for re-processing metadata records into structured and semantic-enriched collection in specific domains, as a platform for the searching over heterogeneous collections, and as a source of inspiration for exploiting ontologies in information searching.

1. RESEARCH QUESTIONS

Searching over heterogeneous metadata records is a notorious problem in digital library community since a huge number of digital libraries created in various metadata standards exist as of today, ranging from academia to government and industry. The leading motivation of searching heterogeneous repositories is that few digital libraries may be able to provide most of the information user might want, and then a federated search over multiple resources is expected such that more relevant results might be discovered. Meanwhile, different digital libraries may adopt distinct metadata formats according to their own specific demands. Thus, such a federated search method must consider as well the interoperation in heterogeneous records.

Federated digital library (FDL) is one of the classic solutions for sharing information among libraries in relevant topics. According to [2], FDL is described as a group of organizations, working together formally or informally, that agree to support a set of *common services* and standards, thus providing interoperability among their members. Conventionally, most of FDLs are built in client/server fashion. However, with the sharp growth of enormous number of digital libraries, especially those moderate-sized ones, there becomes a performance bottle problem in FDLs. One of the revealing example is the 'failure to search' problem existed in earlier version NCSTRL library [9]. Although NCSTRL is now divided into regions and lookup and search are conducted in local region such that load on the central server is alleviated somehow, but how it could be if the number increases another thousands of times? Actually the increase of another thousands of times in data volume is not unforseeable. Therefore, how to adapt to the requirement in scalability, especially in the situation that enormous number of moderate-sized libraries emerge?

Additionally, questions come not only from the architecture scalability, but also from the granularity and meaning in metadata elements[10]. Due to the semantic heterogeneity resulting from the different metadata schemas employed by various digital libraries, the search itself is a difficult and complex task. For example, when describing bibliographic records, metadata experts can adopt MARC [17] to get rather a precise and professional description, while average users many rely on fifteen simple Dublin Core (DC) [3] elements to annotate their personal collections. We argue here that a *common* and *simple* data model is required to harmonize various metadata records. The common feature is easy to understand since a publicly interpretable data model is a necessity for all participating peers in interchanging or integrating semantically related terms. As to the simplicity, it has also been justified that a few large, complex, consistent ontologies, shared by great numbers of users is hard to be created [13]. Instead, small and simple ontologies are easily to be applied, especially in specific domains. The widely acceptance of Dublin Core in library community seems to make it feasible that it can act as a global schema for the interoperability. However, is Dublin Core complete enough to accommodate most of the metadata elements since too many critical information will be lost in the conversion? So, if not, what perspectives should be improved to improve its adaptation to most metadata?

Furthermore, due to words generally may have more than one sense, it is important to choose right senses in the searching according to the context. Besides, broad, narrow or related terms may also help if no results return for current terms. How to realize such a functionality is also a big challenge in both P2P systems and centralized ones.

To sum up, we look into the research question in three aspects since they are highly coherent. The first issue is to study the network topologies which can cope with the searching in highly distributed network. Next, how to present/organize the local repository into a knowledge way such that semantic search can be conducted? Finally, how to exploit the ontology to rewrite user query and achieve more appropriate results?

The rest of the paper is organized as follows. Section 2 presents the current state of exist solutions. from the three questions mentioned above. The considerations on these solutions are presented in the end of each subsection. Section 3 presents the current approach we have achieved so far and also problems we have met. Section 4 describes the expected contributions to the searching issue in P2P based digital libraries. Section 5 concludes the papers.

2. STATE OF EXIST SOLUTIONS

In this section, we present the state-of-art of exist solutions from three perspectives of system architecture, knowledge representation of local repository and user query rewriting.

2.1 System Architecture

As mentioned above, other models are required instead of client/server model to cope with the enormous increase of data volume. We base our prototype system on Peer-to-Peer (P2P) architecture. P2P architecture has been extensively discussed and advertised recently for its advantages over centralized approaches. The P2P paradigm holds many promises: it scales well, by allowing peers (digital libraries) leave and join freely; it strengthens system robustness, with less vulnerability to failure of a central node; it provides federated search across peers of common interest which may contain records not crawled in the 'Hidden Web'.

In P2P network, federated search can be conducted over other connected peers through a local peer, and meanwhile the local peer itself can act as a data provider to make local information accessible to others.

A brief summary of popular P2P systems is described Table 1. The summaries are not complete but collect distinguished features which should be considered in building prospective peer-based FDLs.

2.2 Gnutella, Naspter and Freenet

Gnutella [18], Napster [21] and Freenet [6] are ancestor in P2P computing. They all only support keyword-based search. Gnutella is representative instance for the query flooding which can not scale well. Napster goes to the opposite way and adopt central servers to maintain a centralized directory from which connected peers can register their expertises and also retrieve a list of peers of users' interest. In Freenet, each file/document is identified by a binary key which is generated using some hash function; each peer maintains a local routing table which keeps information about neighbouring peers and the keys are a sequence of *(file key, node address)* pairs used for retrieval. Lookups in Freenet take the form of searches for cached copies. This allows Freenet to provide a degree of anonymity, but prevents it from guaranteeing retrieval of existing documents or from providing low bounds on retrieval costs.

2.2.1 Routing Indices

Crespo [7] uses Routing Indices (RIs) to forward queries to neighbours that are more likely to have answers. If a neighbouring peer cannot answer a query, it forwards the query to a subset of its neighbours, based on its local RI, rather than by selecting neighbours at random or by flooding the network by forwarding the query to all neighbors. RIs are created and maintained by each peers, thus any updates on RI, such as peer joining or leaving, lead to a cascade of updates in peer network. It is also the overhead RIs generated for the sake of efficiency. Such systems also specify relations between query 'topics' and neighbouring peers, which moves further than pure keyword-based query.

2.2.2 Chord, CAN

Distributed hash table (DHT) is probably the most widely used algorithm in P2P computing. Chord [14], CAN [26] are two representative systems using DHT. DHT specifies a relation between entities (files, documents, and etc.) and an identifier (id) in P2P network. Each peer has a unique node ID The value is stored at the node whose ID is closest to the key which is generated by the hash function. The storage and lookups of values are distributed among multiple hosts. Performance of all DHT algorithms has been justified pretty good. For instances, each node maintains information only about $O(logN)$ other nodes, and a lookup requires $O(logN)$ messages [15]. DHT is probably the most applicable algorithm in current research.

2.2.3 pSearch

The major feature of pSearch [28] is to combine the efficiency of DHT systems and accuracy of information retrieval algorithms. [29] also reports a perceivable improvement can be gained in logical routing cost by adopting the proposed algorithms. One weakness in adopting this architecture to digital library is that uneven distribution of document pointers over the peers, and the global state are required beforehand to support the algorithm used in pSearch.

2.2.4 Piazza

Piazza is a peer data management system that enables sharing heterogeneous data in a distributed and scalable way. Piazza assumes the participants have similar content to share within other peers, then makes it possible to define pair wise mappings between their schemas. Then, users formulate queries over their preferred schema, and a query answering system expands recursively any mappings relevant to the query, retrieving data from other peers. Actually, the global as view (GaV) and local as view (LaV), or even both as view (BaV) based methods all have their most suitable application scenarios. We will come to this point in next section.

2.2.5 JXTA Search/JXTA, Edutella, Bibster

JXTA [23] is a P2P interoperability framework created by Sun Microsystems. Peers in JXTA framework are allowed to advertise its own resources through the Peer Discovery Protocol (PDP), then other peers are able to discover such resources. Specifically, each peer is described and published using an advertisement, which is an XML document describing a network resource. Both of Edutella [22] Bibster [12] are built on JXTA framework and aim to combine metadata with P2P networks. They focus on educational domain and bibliographic records respectively, though. Edutella uses super-peer based topologies, in which peers are organized in hypercubes to route queries. In contrast, Bibster is embedded in the general SWAP [25] [8] architecture and a running application.

2.2.6 HyperCup

HyperCup [27] proposes a graph topology which allows for very efficient broadcast and search which intend to reach all peers in

Table 1: Typical P2P Systems Summary

System	Markup-Scheme	Hash Table Usage	Semantic Routing	Query Forwarding	Semantic Query
Gnutella	Keyword	No	No	Yes	No
Naspter	Keywrod	No	No	No	No
Freenet	Keywrod	Yes(binary)	Serial	Yes	No
Routing Indices	Keyword	No	Serial	Yes	No
Chord	Keyword	Yes	Parallel	Yes	No
CAN	Keyword	Yes	Parallel	Yes	No
pSearch	Keyword	Yes	No	Yes[1]	No
Piazza	Database	No	No	Yes	Yes
JXTA Search	XML	No	Parallel	Yes	No
Edutella	RDF	No	Parallel	Yes	Yes (regional)
Bibster	RDF/DAML+OIL	No	Parallel	Yes	Yes (global)
OAI-P2P(ongoing)	RDF	No	Parallel	Yes	Yes (regional)
RDFPeers	RDF	Yes	Parallel	Yes	Yes (global)
HyperCup	Keyword	Yes	Separate Hyper-Cube	Yes	Yes

the network with the minimum number of messages possible. The number of messages generated when peers leave and join the network is $O(log_b N)$ (b is the base of the hypercube), which can be more efficient than DHT algorithm. Moreover, a global ontology is proposed to determine the organization of peers in the graph topology, allowing for efficient concept-based search. This project is rather suitable in our scenario in selecting relevant peers.

2.2.7 RDFPeers

RDFPeers [4] is a scalable distributed RDF repository that stores each triple at three places in a multi-attribute addressable network by applying globally known hash functions to its subject, predicate, and object. Such approach is more suitable to search through highly distributed RDF repositories.

2.2.8 Discussion

Keyword based P2P systems are sufficient for applications which do not need complex query languages nor complex metadata, such as sharing MP3 files. However, for educational resources in Edutella, or bibliographic records in Bibster which may cover title, author, booktitle, etc, queries are more complex and have to build upon standards like IEEE-LOM/IMS [19] or Bibtex metadata with up to 100 metadata entries, which might even be complemented by specific extensions, such as a global ontology is needed to integrate heterogeneous records if they are described in distinct metadata formats. The ordinary information retrieval method is not able to reflect the desired metadata of the creator, title, etc, since it is simply a statistical operation. Evidently, complex forms of metadata make greater storage demands on each individual peer, as each peer must somehow maintain records of more complex markup. However, more detailed metadata seems to be essential if complex queries are to be supported. Edutella and Bibster indicate such issue and demonstrate the possibility to conduct complex queries over metadata records. However, none of them consider harmonizing heterogeneous but related metadata records. The ongoing OAI-P2P project [1] aims to design a P2P network for open archives, where data providers form a P2P network which supports distributed search over all connected metadata repositories. In this scenario, service providers can be removed from this network and make the data repositories more up-to-date. RDFPeers extends the DHT to support the search over RDF triples but it is insufficient to conduct effective and efficient semantic based search since triples are highly distributed and no predetermined RDF Schema is used to support inferencing. In summary, the JXTA framework seems to be more suitable in constructing federated digital libraries since such 'advertisement' can be adopted in selecting appropriate peers ('libraries') with the prerequisite that all peers do advertise their resources/expertises in JXTA network. The comparably more JXTA-based applications can also give us more references.

2.3 Common Data Model

There are already some proposes exist, such as FRBR, CIDOC/CRM, ABC ontology and DC/INDECS. which can be regarded as upper level ontologies in digital libraries, such as

2.4 Exploit Ontologies for Query Expansion

3. PEER SELECTION

Generally, P2P systems are distributed systems without any centralized control, where currently few mechanism has been applied to select suitable peers so as to reduce the blind broadcasting. Haase [11] first indicated such concern recently. The author states that two core operations in most P2P systems are (1) finding the right peers fro querying and (2) efficient routing of message. The first operation in our concern is how to discover/select suitable peers which is able to provide answers to a given query. [11] uses a shared ontology (ACM hierarchical topics for computer science) to advertise peers' expertise in P2P network. In this way, the knowledge about the expertise of other peers forms a semantic topology, independent of the underlying network topology. Although it only considers a rather limited domain (computer science), it is feasible to extend it to a certain classification system, which is widely used in digital libraries, such as Dewy, UDC and LLC, etc. A upgraded query procedure in JXTA framework is depicted in Figure 1, while generally peer just broadcasts the query inside peer group it belongs to.

In Figure 1, when a query is generated in peer A, peer A broadcasts a formatted and concise requests to retrieve other peers' expertises; then A match those expertises with the conducted query automatically and select the suitable peers to send the query to. Such mechanism relieves the load in bandwidth in broadcasting the query and returning all results from other peers in local peer group. Moreover, it saves useless computing resources on peers which can not pro-

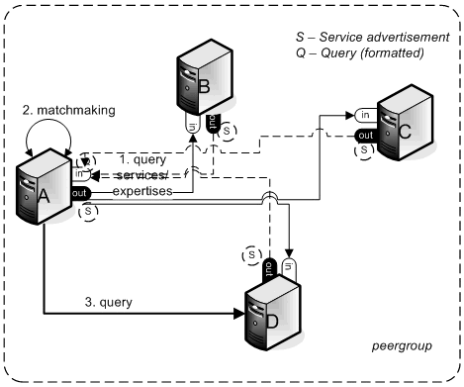

Figure 1: The Upgraded Query Procedure in JXTA Framework

vide any suitable results. The advertisement is used as the primary tool for making general peer, peer group, and resource configuration information available to the network, peers, and peer groups. We can use XML, which is default format in JXTA, to pass advertisement from peer to peer, easily providing expandable and hierarchical representation of information which is compatible with any classification systems.

4. METADATA/ONTOLOGY INTEGRATION

Before we come to discuss the metadata/ontology integration mechanism, it is necessary to differentiate the meanings of 'metadata' in the context of P2P network and digital library. In digital library, metadata has been understood as 'Data about Data'. For example, a library catalogue contains information (metadata) about publications (data), and a file system maintains permissions (metadata) about files (data). Specifically, metadata in digital libraries is the descriptive information of digital objects and it is used to provide user a more efficient way to understand the content of digital object.

However, in P2P network, 'metadata' is adapted to have distinct meaning in different situations. In order to differentiate the usage of the term - metadata, it is necessary to clarify context which are illustrated in following table.

Table 2: 'Metadata' Type in P2P Network

'Metadata' Type	Description
File ID	1. arbitrarily assigned by system according to some mechanism 2. generated by hash algorithm according to practical documents
Keyword ID	keyword is generated by IR method(i.e., TFIDF) and each keyword is assigned an id.
Tuple	a semantic based tuple, such as a RDF statement.

The main difference between the usages of metadata in P2P network and digital libraries is that data in each node in P2P network is given a unique identifier that can rather improve the searching quality by adopting corresponding algorithms, while the metadata in digital libraries is a set of extra descriptive information on the data and it is not mandatory that metadata information must be embedded in the data. In order not mix them up, we in this paper use 'ontology' to indicate metadata records in digital library since they can be viewed as ontologies to certain degree. Meanwhile, we still keep the term 'metadata' to indicate the identifier information in P2P network.

The second main goal of this project is to make it easier to search in peer based federated digital libraries, and more peculiarly, to search across those collections created in heterogeneous metadata sets/domain ontologies. It is explicitly designated with the adaptations of the Semantic Web technologies to deal with the heterogeneous metadata. To make it simple, we do not discuss the syntax and structure issues about the repositories here and we consequently assume that all recordsets created in varied metadata are XML formatted. In order to explicate the semantic meanings and deduce over the relationships among these records, we should encode the records in RDF or OWL. Here, we argue that RDF/RDFS is used as the format of repository instead of OWL. Generally, OWL allows the definition of specific constraints on the data model, and has more expressive power, such as symmetric, inverse and transitive (i.e. OWL Full). However, RDFS allows greater flexibility, and this may be an advantage. For example, OWL Lite and OWL DL do not allow a class to be also an instance, but this kind of meta-modelling may be useful for thesaurus data. In addition, tool support for RDFS is perhaps at this stage more mature.

Moving further to the ontology integration concern in P2P network, from the theoretical perspective, both of the global as view (GaV) method and the local as view (LaV) based method are applicable in querying heterogeneous records. (We will not discuss the basic definitions of GaV and LaV and refer interested readers to [16].) However, LaV approach is considered to be more difficult. In fact, the only knowledge we have about the data in the global schema is through the views representing the sources, and such views in LaV provide only partial information about the data. As to the GaV approach, query answering is just similar to a simple unfolding strategy since it maps directly sources queries to corresponding elements in the global schema [16].

Actually, many of the work carried out on ontologies integration for the semantic Web is on to build the global ontology on the basis of the local ones [5]. We can adopt GaV based method as well and take MODS (Metadata Object Description Schema) [24] as the global ontology, which is a more expressive alternative to Dublin Core but still keep the simplicity which make it easier to adopt in representing digital objects. Obviously, if one has two relevant ontologies, it is not difficult to convert a more complex ontology to a simpler one. Some information may be lost or granularity may become coarse, though. However, we may feel it out of the question in a reverse way. In subtotal, MODS has 19 top level elements and 47 sub elements, and is organized in a hierarchical structure. According to the introduction in MODS' official website, some advantages of MODS are:

- The element set is richer than Dublin Core
- The element set is more compatible with library data than ONIX
- The schema is more end user oriented than the full MARCXML schema
- The element set is simpler than the full MARC format

In our prototype, all the local collections, which may be created in distinct ontology, are to be transformed into an global ontology based knowledge repository.

The relations between source ontology and target ontology (MODS) should be explicit during the conversion. Since many of the relationships between ontology elements are not only equivalent, there may have narrower terms, broader terms or related terms. Once there is a possibility to query the provenance information, such explicit mapping information will be of help in better understanding.

To achieve the aim mentioned in last section, a knowledge organizing language is required to delineate relationships within concepts and make them interpretable. Hereby, we propose to use SKOS (Simple Knowledge Organization Systems) [20] to set up the relationships among local ontology and MODS.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented the critical issues in designing and implementing peer based federated digital libraries. Two major goals in this project have been presented. One is to find peers which are able to answering user queries in a more effective and efficient way; the other is to support semantic search cross distributed and heterogeneous repositories. We focused on three major concerns which are rather critical in constructing semantic based FDL in P2P network, namely, platforms election, peer selection and metadata integration. Generally, there are interesting alternatives for each of the different aspects and , actually, we are still exploring multiple approaches to optimize the overall system. In practice, however, it constitutes a major challenge to integrate these concerns together to realize such a peer based FDLs.

The future work in implementing our system is consisted of: (1)finish the prototype work (e.g., integrate semantic similarity function for peer selection component); (2) investigate more repositories for evaluation; (3) system optimization (e.g., query optimization).

6. REFERENCES

[1] AHLBORN, B., NEJDL, W., AND SIBERSKI, W. OAI-P2P: A Peer-to-Peer Network for Open Archives . In *2002 International Conference on Parallel Processing Workshops (ICPPW'02)* (Vancouver, B.C., Canada, 2002).

[2] ARMS, W. Y. *Digital Libraries*. The MIT Press, 2000.

[3] BOARD, D. U. Dublin core medata initiative. http://www.dublincore.org, 2003. Available: 2004.07.

[4] CAI, M., AND FRANK, M. Rdfpeers: A scalable distributed rdf repository based on a structured peer-to-peer network. In *The 13th International World Wide Web Conference (WWW)* (2004), pp. 17–22.

[5] CALVANESE, D., GIACOMO, G. D., AND LENZERINI, M. A Framework for Ontology Integration. In *International Semantic Web Working Symposium(SWWS)* (Stanford University, California, USA, 2001).

[6] CLARKE, I., SANDBERG, O., WILEY, B., AND HONG, T. W. Freenet: A distributed anonymous information storage and retrieval system. In *International Workshop on Design Issues in Anonymity and Unobservability, LNCS 2009* (Berkeley, CA, USA, 2000), vol. 2009 of *Lecture Notes in Computer Science*, Springer, pp. 46–66.

[7] CRESPO, A., AND GARCIA-MOLINA, H. Routing indices for peer-to-peer systems. In *Proceedings of the 22 nd International Conference on Distributed Computing Systems (ICDCS'02)* (2002).

[8] EHRIG, M., HAASE, P., SIEBES, R., STAAB, S., STUCKENSCHMIDT, H., STUDER, R., AND TEMPICH, C. Invited contribution the swap data and metadata model for semantics-based peer-to-peer systems. In *MATES* (2003), vol. 2831 of *Lecture Notes in Computer Science*, Springer, pp. 144–155.

[9] FOX, E., FRENCH, J., JONES, P., MALY, K., NELSON, M., AND ZUBAIR, M. NCSTRL - Networked Computer Science Technical Reference Library. http://www.ncstrl.org/, Available: 2005.03.

[10] GODBY, C. J., YOUNG, J. A., AND CHILDRESS, E. A repository of metadata crosswalks. *D-Lib Magazine 10*, 12 (2004).

[11] HAASE, P., AND SIEBES, R. Peer selection in peertopeer networks with semantic topologies. In *The 13th International World Wide Web Conference (WWW)* (New York, NY USA., May 17-22, 2004).

[12] HAASE1, P., BROEKSTRA3, J., EHRIG1, M., MENKEN, M., AND ETC. Bibster — a semantics-based bibliographic peer-to-peer system. In *The 3rd International Semantic Web Conference (ISWC2004)* (Japan, 2004).

[13] HENDLER, J. A. Agents and the semantic web. *IEEE Intelligent Systems 16(2)* (2001), 30–37.

[14] ION STOICA, ROBERT MORRIS, D. K., KAASHOEK, F., AND BALAKRISHNAN, H. Chold: A scalable peer-to-peer lookup service for internet applications. In *SIGCOMM* (San Diego, California, USA, 2001).

[15] JAIN, S., MAHAJAN, R., AND WETHERALL, D. A Study of the Performance Potential of DHT-based Overlays. In *USENIX Symposium on Internet Technologies and Systems 2003* (Seattle, Washington, USA, 2003).

[16] LENZERINI, M. Data integration: A theoretical perspective. In *Proceedings of the Twenty-first ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems(PODS)* (Madison, Wisconsin, USA, 2002), pp. 233–246.

[17] LIBRARY OF CONGRESS. Marc standards. (n.d.). library of congress network development and marc standards office. http://lcweb.loc.gov/marc/, Available: 2004.07.

[18] LIME WIRE LLC. Gnutella - limewire 4.0. http://www.limewire.com/, 2000.

[19] LTSC-LOM. Draft standard for learning object metadata. http://ltsc.ieee.org/wg12/, 2002.

[20] MILES, A. J., ROGERS, N., AND BECKETT, D. W3c an rdf schema for thesauri (skos-core 1.0 guide). http://www.w3.org/2001/sw/Europe/reports/thes/8.1/, November 2004.

[21] NAPSTER, LLC. Napster. http://www.napster.com, 2001.

[22] NEJDL, W., WOLF, B., QU, C., DECKER, S., SINTEK, M., NAEVE, A., NILSSON, M., PALMER, M., AND RISCH, T. EDUTELLA: a P2P networking infrastructure based on RDF. In *International World Wide Web Conferences (WWW)* (2002), pp. 604–615.

[23] OAKS, S., TRAVERSAT, B., AND GONG, L. *JXTA in a Nutshell*. O'Reilly & Associates, Inc., September, 2002.

[24] OF CONGRESS, T. L. Metadata Object Description Schema(MODS). http://www.loc.gov/standards/mods/, Available at:2005.03.

[25] PINTO, H. S., STAAB, S., SURE, Y., AND TEMPICH, C. OntoEdit Empowering SWAP: a Case Study in Supporting DIstributed, Loosely-Controlled and evolvInG Engineering of oNTologies (DILIGENT). In *ESWS* (2004), vol. 3053 of *Lecture Notes in Computer Science*, Springer, pp. 16–30.

[26] RATNASAMY, S., FRANCIS, P., HANDLEY, M., KARP1, R., AND SHENKER, S. A scalable content-addressable network. In *SIGCOMM* (San Diego, California, USA, 2001).

[27] SCHLOSSER, M. T., SINTEK, M., DECKER, S., AND NEJDL, W. Hypercup - hypercubes, ontologies, and efficient search on peer-to-peer networks. In *AP2PC* (2002), vol. 2530 of *Lecture Notes in Computer Science*, Springer, pp. 112–124.

[28] TANG, C., XU, Z., AND MAHALINGAM, M. psearch: Information retrieval in structured overlays. In *Proceedings of HotNetsI02* (Princeton, New Jersey, USA, 2002).

[29] XU, Z., TANG, C., AND ZHANG, Z. Building topology-aware overlays using global soft-state. In *Proceedings of the 23rd International Conference on Distributed Computing Systems (ICDCS'03)* (2002).