

Exploring JXTASearch for P2P Educational Media Discovery

Changtao Qu
Learning Lab Lower Saxony
University of Hannover
Expo Plaza 1, D-30539, Hannover, Germany
qu@learninglab.de

Wolfgang Nejdl
Computer Science Department
Stanford University, Stanford, CA 94305, USA
nejdl@db.stanford.edu
(On leave from University of Hannover)

Abstract

In this paper we present a Peer-to-Peer (P2P) application that aims at improving the discoverability of educational media typically distributed over different universities and institutions. We investigate JXTASearch, a distributed search engine constructed upon Sun's open source P2P platform JXTA, and further extend it to support Dublin Core Metadata Element Set (DCMES) based P2P searches. While the JXTA platform provides the essential interoperability among various peers at the "network" level, DCMES constitutes the minimal interoperable basis for educational media discovery at the "metadata" level. We then discuss the advantages and shortcomings of our prototype implementation, as well as some future work building upon our experience with the prototype.

1. Introduction

Although the term of educational media may have a broad definition covering both print media and electronic media with the educational purpose, we refer to educational media in this paper principally as a large number of learning resources that directly exist on the Web and can be obtained by educators and learners in that all learning resources are annotated with corresponding learning resource metadata. These metadata are typically managed in various learning resource metadata repositories distributed over different universities and institutions, which can enable learning resources to be searched and discovered. Generally speaking, in E-Learning these metadata repositories are usually constructed based on several popular learning resource metadata sets, e.g., DCMES (Dublin Core Metadata Element Set) [6], IEEE LOM (Learning Object Metadata)/IMS Learning Resource Metadata Specification [11][12], and ADL (Advanced Distributed Learning) SCORM (Sharable Content Object Reference Model)[1].

In recent years, central-server based approaches to educational media discovery have received lots of attention and have also found many application cases [5][17][20][21]. However, since many universities and institutions are generally reluctant to give up their control over learning resources produced or accumulated by themselves, most of central-server based applications, e.g., E-Learning "portals"[20][21], have run into increasing predicaments. In general, these applications come across more and more difficulties during expanding process, and in most cases are costly but unprofitable. P2P (Peer-to-Peer) just comes into the picture in E-Learning by virtue of its capability of enabling institutions to participate in a global sharing network without losing the original control over their learning resources.

In comparison to central-server based approaches, P2P networks provide some novel techniques for distributed resource discovery, principally represented by several successful P2P search engines, e.g., Napster [14], Gnutella [9], and Freenet [8], etc. Since these P2P applications are specifically designed for distributed file sharing rather than for metadata-centred distributed searches, they can only support resource discovery based on some simple metadata definitions and thus cannot be directly applied for educational media discovery, where rich types of learning resources and diverse sets of learning resource metadata specifications co-exist. Recently, there emerged a metadata-centred distributed P2P search engine JXTASearch [22], which provides us with new opportunities to accomplish efficient P2P educational media discovery.

JXTASearch [13] is a sub-project of Sun's open source project JXTA [18], which is originally designed to realize both "wide" (the ability to access various devices on the Web) and "deep" (the ability to access various backend repositories) searches based on certain metadata sets [22]. It is of importance for educational media discovery mainly due to its two abilities:

First, the ability to provide widespread interoperability among various peers.

As a distributed P2P search engine built upon Sun's JXTA platform, JXTASearch directly inherits JXTA's essential interoperability among various peers. Regarding content provider peers, JXTASearch is able to not only work well with pure JXTA provider peers, but can also bring Web-based provider peers into the searching scope without great overhead. This ability facilitates the shaping of a content provider scenario for educational media discovery, where heterogeneous learning resource metadata repositories co-exist. In addition, regarding content consumer peers, JXTASearch can integrate pure JXTA peers, Web peers, and some low-end distributed devices, e.g., PDAs, handhelds, and cell phones, etc., into the JXTASearch network. Generally, at the "network" level, the JXTA platform basically ensures a widespread searching scope and potentially large user community of JXTASearch.

Second, the ability to provide good adaptability to various metadata sets.

In P2P networks, using metadata is almost the only way for a peer to expose its own resources or discover and retrieve shared resources from other peers. While a simple metadata definition is sufficient for some domain-specific P2P search engines, e.g., Napster, Gnutella, etc., discovering educational media in P2P environments requires more complex metadata support, having to involve in most cases several advanced learning resource metadata sets, e.g., DCMES, IEEE LOM/IMS, and SCORM. At the "metadata" level, JXTASearch outmatches most of other P2P search engines through introducing a novel searching concept: the queryspace, together also with an open design of its searching algorithm. Through utilizing queryspaces, JXTASearch can be extended to adapt to various metadata sets. Also the channel-like searching approach enabled by queryspaces can greatly improve the searching efficiency in complicated P2P environments.

As the current JXTASearch implementation realizes only one simple and fixed queryspace (though the general design of JXTASearch is quite flexible), we've modified it

by implementing a DCMES queryspace as the first attempt to extend JXTASearch's searching capability. Since DCMES, especially its RDF (Resource Description Framework)/XML (eXtensible Markup Language) binding [2], is the minimal interoperable basis of several other learning resource metadata sets, such as LOM/IMS and SCORM, the "extended" JXTASearch can also handle learning resource repositories, which contain the learning resource metadata of LOM/IMS and SCORM XML binding, taking advantage of a simple mapping approach [16]. Moreover, by means of the extension, the whole JXTASearch "sub-network", including all its underlying provider peers, can be integrated into Edutella [7][15], a RDF-based E-Learning P2P network, through using an Edutella/JXTASearch gateway.

2. General design

2.1. JXTA and JXTASearch

Project JXTA is Sun's P2P initiative focused on constructing an interoperable P2P network [10]. The JXTA platform employs a layered approach, including the JXTA core layer, the JXTA service layer, and the JXTA application layer. At its basis, JXTA is nothing more than six XML-based protocols defined at a much lower level in comparison to some other P2P platforms and applications, which can provide widespread interoperability among various peers. These six protocols are implemented at the JXTA core layer in the form of several building blocks that can be used by almost all P2P applications regardless of their intended users, platforms, devices, and specific implementations. Upon the JXTA core layer we have the JXTA service layer and the JXTA application layer, where some higher-level interoperable services and applications are built. In fact, JXTASearch is designed according to JXTA's 3-layer approach. In figure 1 we illustrate four types of JXTASearch services and three types of JXTASearch applications, as well as their corresponding locations on the layered JXTA platform.

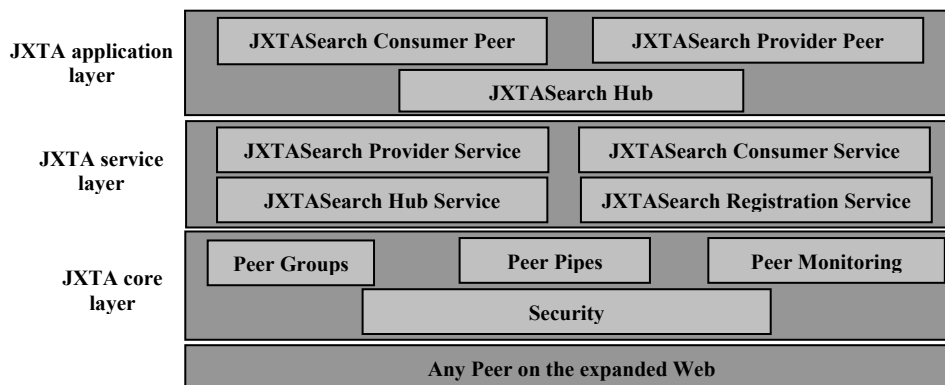


Figure 1. JXTASearch and its location on the layered JXTA platform

The kernel of JXTASearch services is QRP (Query Routing Protocol)[13], which is an XML protocol and fully specifies the message types, message formats, and message routing rules that must be supported by any JXTASearch peer. Based on the QRP, JXTASearch defines four types of services [22]:

JXTASearch Provider Service: a service that accepts queries written in the QRP from JXTASearch hub (or JXTASearch consumer peer directly) and responds in the QRP to the requestor. The Provider Service does not perform any indexing or searching internally, but rather calls the appropriate interface on the provider itself, e.g., database search, etc.

JXTASearch Consumer Service: a service that sends queries written in the QRP to the hub (or to JXTASearch provider peer directly) and awaits QRP responses. The Consumer Service also performs formatting of the responses for presentation to the end user or application.

JXTASearch Registration Service: a service that sends requests for registration to JXTASearch hub and maintains the registration file for the provider.

JXTASearch Hub Service: a service that performs routing of queries from consumers to providers. The Hub Service accepts queries, resolves those queries to the appropriate providers, manages the routing of the queries to the providers, and sends collated results back to the requesting consumer. In general, the Hub Service consists of two sub-components: (1) JXTASearch Router, which routes and manages query connections, collates results and returns results to consumers; and (2) JXTASearch Resolver, which matches queries to providers using a full text search engine that indexes metadata specified by the provider during registration.

At the JXTA application layer there are three types of JXTASearch peers, respectively JXTASearch providers, consumers, and hubs. Because the QRP has two transport bindings respectively for JXTA platform (XML over JXTA using either HTTP or TCP) and the Web (XML over HTTP), JXTASearch providers can be further divided into two sub-types, either pure JXTA peers or Web servers with the JXTASearch adapter. Likewise, JXTASearch consumers also have two sub-types, either pure JXTA peers or Web browsers with HTTP client interface to the JXTASearch network. In figure 2 we illustrate the general architecture of JXTASearch, which describes the interaction between JXTASearch peers and different JXTASearch services [22].

In the JXTASearch network, each JXTA peer can run instances of the JXTASearch Provider Service, Consumer Service, and Registration Service on top of its JXTA core. Each peer interacts with the JXTASearch Hub Service, which itself also runs on top of the JXTA core. In section 3 we'll further describe interactions between different JXTASearch peers along with an introduction to our prototype implementation.

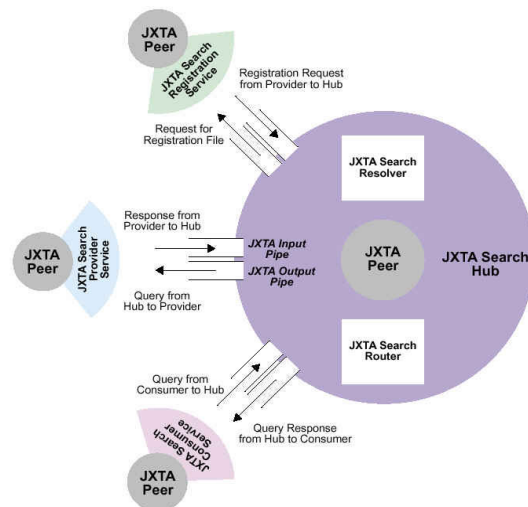


Figure 2. General architecture of JXTASearch [22]

2.2. DCMES and its RDF/XML binding

DCMES is a simple yet effective metadata set for describing a wide range of networked resources. Since the original JXTASearch only defines one queryspace (<http://search.jxta.org/text>), which enables a full-text-like distributed searches, extending JXTASearch with DCMES can refine the searching scope so as to improve the searching efficiency. Since JXTASearch adopts an open design style, the core of JXTASearch, e.g., query resolving and routing algorithm, etc., does not need to be touched during the extension process. We need only to modify part of JXTASearch source code in order to adapt it to the DCMES metadata schema and then construct a new DCMES queryspace (<http://www.learninglab.de/DC>) that defines DCMES-conformant query structure and its associated registration. Within the JXTASearch network, various queryspaces can co-exist, which constitutes a channel-like searching strategy.

In the DCMES queryspace, the DCMES metadata set is represented using its RDF/XML binding [2]. An important consideration here is that the RDF/XML bindings of several other learning resource metadata sets, e.g., LOM/IMS, and SCORM, are defined in a consistent fashion with DCMES RDF/XML binding. Therefore, using DCMES RDF/XML binding to construct the new queryspace will facilitate the future integration of LOM/IMS and SCORM based learning resource metadata repositories into the JXTASearch network (further also the RDF-based Edutella network [16]).

3. Prototype implementation

In figure 3 we illustrate the system architecture of the prototype implementation.


```

<query><dc:publisher/></query>
<query><dc:contributor/></query>
<query><dc:date/></query>
<query><dc:type> Dataset Image etc.</dc:type></query>
<query><dc:format>text/html etc.</dc:format></query>
<query><dc:identifier/></query>
<query><dc:source/></query>
<query><dc:language>de en</dc:language></query>
<query><dc:relation/></query>
<query><dc:coverage/></query>
<query><dc:rights/></query>
</or>
</predicate>
</query-space>
</register>

```

Figure 4. A simplified registration file of a JXTA provider peer

3.3. Searching and discovering

Similar to JXTASearch provider peers, there are two types of consumer peers in the JXTASearch network: either pure JXTA consumer peers or Web consumer peers (usually Web browsers) with the JXTASearch interface. The essential difference between the both is that JXTA consumer peers can directly send queries to any JXTA provider peers (including JXTASearch hubs) and then get direct response from these peers, whereas Web consumer peers must use JXTASearch hubs as mediators in order to accomplish searching and discovering. In the prototype implementation, we simply use an extended JXTA Shell [19] as a JXTA consumer peer, and use a JSP (Java Server Pages) running on JXTASearch hubs as the JXTASearch interface for Web consumer peers. In figure 5 we illustrate the query GUI of a Web consumer peer.

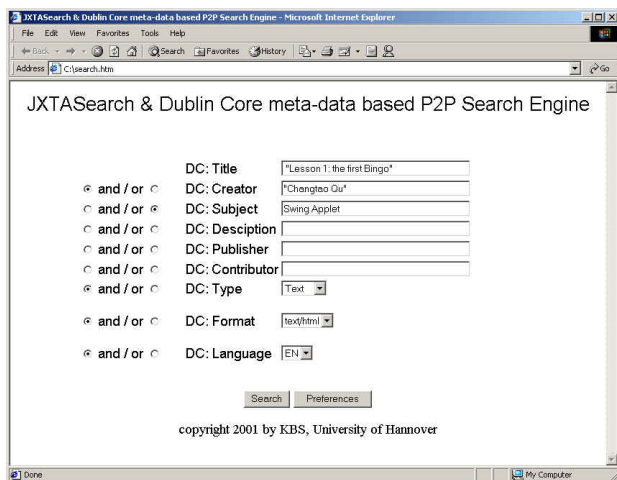


Figure 5. The query GUI of a Web consumer peer

From DCMES' 15 metadata definitions, we select nine commonly used metadata entries in order to format a query. Also users can define logical relationships between

selected metadata entries. Note that these logical relationships only make sense for local query services, they do not have influence on the query resolving and routing process.

After discovering, the query results formatted in XML syntax according to the QRP can be rendered by consumer peers in different styles. For Web consumer peers, JXTASearch hub (precisely speaking, hub's HTTP Router) is responsible for rendering the results in a table-like style and sending them back to the consumers. For JXTA consumer peers, the JXTASearch Consumer Service running on each consumer peer is responsible for presenting the query results. In our prototype implementation, JXTA Shell simply renders the query results into formatted texts.

3.4. Resolving and routing

The query resolving is the process of determining to which set of providers a given query should be routed. The minimal condition for sending a query to a provider peer in the JXTASearch network is that the query must have the same queryspace as the provider registration. In JXTASearch, query resolving is implemented by a Java Servlet (JXTASearch Resolver) running on the hubs. It can be directly called by JXTASearch Routers.

In general, the JXTASearch Router sequentially performs three functions: (1) receive queries from consumer peers; (2) call JXTASearch Resolver to find appropriate providers and further rout the queries to the providers; and (3) merge query results and send them back to consumer peers. Corresponding to JXTASearch's two transport bindings, there are two types of JXTASearch Routers: (1) JXTA Router, used to handle queries from JXTA consumer peers utilizing pipes; and (2) HTTP Router, used to handle queries from Web consumer peers utilizing HTTP. Because the query resolving and routing algorithm as well as their implementations in JXTASearch are shielded from end-user applications, they do not need to be touched during the prototype implementation.

4. Discussion

In comparison to conventional crawler-based search engines, distributed P2P search engines such as JXTASearch have many notable advantages, e.g., the ability to provide ad hoc query results, the ability to cross firewalls and heterogeneous backend repositories, as well as high searching efficiency enabled by some P2P-specific designs, such as JXTA's "peer group", and the queryspace proposed by JXTASearch. Together with the powerful metadata support, these P2P search engines might essentially change the way information is shared, which makes special sense for the sharing of educational media in E-Learning.

Our early exploration to JXTASearch also manifested several shortcomings of JXTASearch, one of which is concerned with the QRP, which defines simple query, response, and registration formats using XML syntax. For more complex queries, the current QRP is insufficient, and would have to be extended along the lines of current XML query languages, which will unfortunately generate damage to the current JXTASearch core implementation, e.g., the query resolving and routing algorithm, etc.

Another shortcoming of JXTASearch lies in its missing of multi-hub chaining. Although JXTASearch does contain mechanisms for supporting multiple hubs, the JXTASearch protocol [13] does not yet define methods for chaining hubs. In fact, coordination among a number of hubs enabled by some sophisticated mechanisms is one key to ensuring the widespread searching scope and high searching efficiency of JXTASearch. This issue has to be addressed by future JXTASearch protocol and JXTASearch hub implementation.

5. Conclusions and future work

Using JXTASearch and DCMES, our prototype implementation explored the possibility of building a P2P framework that can be applied for improving the discoverability of educational media. Based on the work described in this paper, we are now working on a more advanced P2P framework Edutella, a RDF-based P2P infrastructure for the exchange of educational media in the global area [7][15]. Edutella extends our JXTASearch prototype by introducing the design of more general hubs and especially a more powerful query language (a Datalog-based query language transmitted in RDF/XML syntax), which can express not only simple conjunctive queries as defined by JXTASearch QRP, but also relationally complete queries based on RDF namespaces to query arbitrary schemas over the net. As an official JXTA service, Edutella will co-exist with JXTASearch on the future JXTA platform, interfacing with JXTASearch through an Edutella/JXTASearch gateway that we are currently working on.

6. References

- [1] ADL Technical Team, SCORM Specification V1.2, <http://www.adlnet.org/index.cfm?fuseaction=scormabt>
- [2] Beckett, D., E. Miller, and D. Brickley, Expressing Simple Dublin Core in RDF/XML, <http://dublincore.org/documents/2001/09/20/dcmes-xml/>
- [3] Clark, J., and S. DeRose, XML Path Language (XPath), <http://www.w3.org/TR/xpath>
- [4] dbXML Group, dbXML Native Database, <http://www.dbXML.org/>
- [5] Dhraief, H., W. Nejdl, and B. Wolf, Open Learning Repositories and Metadata Modeling, in Proc. of International Semantic Web Working Symposium, Stanford, USA, Aug. 2001.
- [6] Dublin Core Metadata Initiative, Dublin Core Metadata Element Set, V 1.1, <http://dublincore.org/documents/1999/07/02/dces/>
- [7] Edutella, <http://edutella.jxta.org>
- [8] Freenet, <http://www.freenetproject.org>
- [9] Gnutella, <http://www.gnutella.com>
- [10] Gong, L., Project JXTA: A Technology Overview, <http://www.jxta.org/project/www/docs/TechOverview.pdf>
- [11] IEEE Learning Technology Standards Committee, IEEE LOM working draft 6.4, <http://ltsc.ieee.org/wg12/index.html>
- [12] IMS Global Learning Consortium, Inc., IMS Learning Resource Metadata Specification V1.2.1, <http://www.imsproject.org/metadata/index.html>
- [13] JXTASearch Engineering Team, JXTA Search Protocol Spec. 1.0 (Draft), <http://search.jxta.org/protocol.html>
- [14] Napster, <http://www.napster.com/>
- [15] Nejdl, W., B. Wolf, C. Qu, S. Decker, M. Stintek, A. Naeve, M. Nilsson, M. Palmer, and T. Risch, Edutella: A P2P Networking Infrastructure Based on RDF, in Proc. of the 11th International World Wide Web Conference (WWW 2002), Hawaii, USA, May 2002.
- [16] Qu, C., W. Nejdl, and H. Schinzel, Integrating Schema-specific Native XML Repositories into a RDF-based E-Learning P2P Network, Technical Report, Learning Lab Lower Saxony, July 2002.
- [17] Open Archives Initiative, <http://www.openarchives.org/>
- [18] Sun Microsystems Inc., JXTA v1.0 Protocols Spec., <http://spec.jxta.org/v1.0/docbook/JXTAProtocols.html>
- [19] Sun Microsystems Inc., Project JXTA: Technical Shell Overview, <http://www.jxta.org/project/www/docs/TechShellOverview.pdf>
- [20] Teachware On Demand, <http://www.teachware-on-demand.de/index.php>
- [21] UNIVERSAL Project, <http://www.ist-universal.org>
- [22] Waterhouse, S., JXTASearch: Distributed Search for Distributed Networks, <http://search.jxta.org/JXTASearch.pdf>