In Proceedings of ED-MEDIA 2003, pages abstract on page 36, (full paper on CD).
Association for the Advancement of Computing in Education (AACE), June 2003.
ISBN 380094-48-7

# Sharing and Search for Learning Materials by Semantically Enriched Peer-to-Peer Networks

Luka Divac-Krnic, Cornelia Seeberg and Ralf Steinmetz

KOM - Multimedia Communications
Department of Electrical Engineering and Information Technology
Darmstadt, University of Technology
Merckstr. 25 • 64283 Darmstadt • Germany

{divac-krnic, seeberg, steinmetz}@kom.tu-darmstadt.de

**Abstract:** *In the absence of the Semantic Web alternatives using peer-to-peer networks may be worked out. A bold assertion is made that peer-to-peer networks can get more efficient and performant than the Semantic Web with all its potentials can be incepted. On the way towards semantically and technically enhanced peer-to-peer networks an overview shall be given which interests shall be targeted and who or what institutions can mostly benefit from such networks considering the resources at disposition. One sure beneficiary from our point of view would be educational bodies like universities which would profit from each other through wider proliferation of learning materials.*

**Keywords:** peer-to-peer networks, semantics, Semantic Web, learning materials

## 1 Introduction

In the course of enhancing peer-to-peer (P2P) networks an adding of semantics to the present generation of P2P-applications that accomplish keyword-related searches would give an impetus to the accuracy of search results. Educational bodies like universities would surely benefit from resulting proliferation of learning contents through such P2P networks.

However the search of resources getting quicker and preciser, it does not mean that the user will be able to **use** the found resources quicker - the download might still be an obstacle. To clarify for what environment such new semantically enhanced P2P-networks would be useful, one has to define a precise scope of usage herein. An educational use as a network of universities exchanging their learning materials makes sense, providing authors easy maintenance of their learning contents that reside only on their hard-drives.

The part 2 of this paper shows the implications of lack of semantics in P2P-networks and its possibilities and those of the Semantic Web. Part 3 displays related works, whereas the part 4 deals with criteria for assessment of future directions for development of P2P-networks. The papers closes with conclusion in part 5.

## 2. The Absence of Semantics in Peer-to-Peer Networks

The motivation of P2P networks today is rather entertainment-based than of educational nature. The search of files is keyword-based and is in case of music-files sharing mostly related to the title of the searched file or the name of the performer, that are incorporated in the title of the wanted file.

But if one is for instance looking for a "crane", the object found at the certain peer could have completely different meanings, because crane is kind of bird and at the same time a machine for lifting objects like building materials. So if you are interested in ornithology and looking for some documents about cranes, you are not expecting to find a company lending cranes for building a house. One clearly needs to have more expressiveness in search requests for content like: "ornithological description of a bird crane" in order to narrow the search on contents that make sense and get accurate hits. The reason herefore is the higher semantic complexity of an unspecified learning resource (like a lecture about cranes) in comparison to e.g. a music file where a title of a song and the name of a performer have a sufficient "semantic richness" to describe the searched object.

### 2.2 Semantic Web

The Semantic Web will be a powerful tool, which will in the words of Tim Berners-Lee [THL2001]: "Allow anyone to say anything about anything". Besides an improved search of documents, it shall be able to facilitate selecting and triggering of services found on the Web or exchanging data with other agents without any human intervention.

In order to make machines perform adequate document searches, one has to provide meaning of the documents to machines. Relational informations between documents need to be specified and tagged to these documents. Documents will be annotated with markup - metadata, that would express their semantic meaning. An additional way of providing semantic expressions of relations between objects is building of ontologies for specific domains of real world. The objects are thus arranged in super- and sub-classes, which could be easily expanded or replaced by new classes of objects.

In the building of the Semantic Web RDF (Resource Description Framework) will play a big role as a foundation for processing metadata [RDF99]. RDF provides expression of "triples", in form of showing relations between objects by supplying a pointed correlation between them. The most popular means of transport of RDF Metadata is XML based and would constitute a supplementary layer on top of XML. Further involved technologies supporting will probably include Topic Maps  XTM [XTM2001], and ontology description language like DAML+OIL [DAO2001] or OWL [OWL2002].

### 2.3 Peer-to-Peer Networks

P2P communication constitutes a model where each participating party has same capabilities, in contrast to client/server paradigm. Its technology facilitates providing of services of any online device to any other online device. The potentials of P2P networks go far beyond popular file sharings, like in the case of Gnutella [GNU2000] and Freenet [FRE2001]. They can distribute the responsibilities and thus the load of providing services among peers in a network enabling distributed computing, as in the case of the SETI project [SET2001].

The nature of a P2P network causes some great disadvantages, like requests being sent not resulting in a response at all, because of peers suddenly disconnecting from the network. Though these might be solved with help of their redundant nature enabling replicating resources among peers. Nevertheless there is still not possible to give performance guarantees in P2P networks.

## 3. Related Work

### 3.1 Edutella

Wolfgang Nejdl et al. describe in [EDU2002] Edutella, an RDF-based metadata infrastructure for P2P applications building on JXTA framework [JXT2001]. Beside the query service as a core service of Edutella, a data model (ECDM) and a query exchange language (RDF-QEL-i) are introduced. Further services are replication (availability and workload balancing), mapping (translation between different metadata), mediation  (reconciliation of conflicting or overlapping information) and annotation service. It focuses on exchange of metadata of learning resources that are generated by schemes like IEEE LOM, IMS and ADL SCORM.

### 3.2 SON - Semantic Overlay Networks

Arturo Crespo and Hector Garcia Molina propose in [SON2002] a clustering of nodes with semantically similar content in Semantic Overlay Networks (SONs). Thus the search efficiency should rise because queries are processed by identifying which SON or SONs are better suited to answer it, whereas SONs unlikely to have the searched content are not bothered. In this way node connections within a controlled collection of peers are structured rather than documents.

### 3.3 An alternative approach with help of ontologies

An enrichment of contents with metadata can be accomplished by using a controlled and specified vocabulary of a metadata-set, like LOM [LOM2002]. We created a LOM editor [Ste2002] which is implemented as a Java application with Xindice [XIN2002] database connection. During creation of new LOM metadata set for a specific content 10 fields can automatically be, depending on one’s business model, filled with values, and 13 other fields are provided for alternative manual filling. The 10 automatically filled fields include information which can be easily be obtained on the fly during editing, like name, size and date of creation.

Provision of an ontology which would represent contents within one repository of a peer (or a group of peers clustered in one) can significantly improve search and administering of contents. There are basically two ways of providing insight and understanding of an ontology to other peers: by a consensus to use same kind of ontology (with the same descriptive language) or by an exchange of different ontologies between peers. The latter approach may

be very difficult because one does not know what data each field in an ontology may contain, precipitating different algorithms for extraction of data. But if one takes into consideration that universities with same or similar educational contents (similar professorship chairs) would want to exchange their contents with each other, they are more likely to decide by a consensus which ontology to use.

As the creation of an ontology needs a great deal of thinking about and modeling one's own resources by classifying them into sub- and super-classes, so that a concept of inheritance and inferencing can be achieved, the actual difficulty is making the authors or the owners of these contents performing such a task. A way has to be found to commit every author to add a term in his ontology every time when updating his or hers repository with new contents. If this task is not done properly then the provided semantics would be inaccurate resulting in other peers getting content they were not looking for. This can be solved by introduction of a reward/penalty scheme for peers. If a peer does not properly update his/her ontology this would cause erroneous semantics which would result in inaccurate search by other peers. The peer providing such "false" content shall be granted lower priority for his downloads, in contrast to an author tagging his contents with correct semantics who will be rewarded by getting higher prioritized download sessions.

## 4. Criteria of Assessment

The postulation of a lower performant substitute for "Semantic Web" is simple: there is no Semantic Web so far. Hence the need of some institutions for exchange of scientific material would have to rely on technologies, that already exist and could be enhanced "faster" than the Semantic Web could be built.

Due to sheer size of the World Wide Web today one can envision the magnitude of work needed to create a semantic layer over the existing net now. That is why an effort could be worth making an attempt to provide a semantic search of objects or documents with technologies that are widely available. A suitable area for beginning would be institutions like universities, that need to exchange academic material like papers, diploma theses or different research projects related material and learning material like lecture notes or lessons with multimedia contents. Another area or group are companies in need of an accurate retrieval of services and products form other companies.

A single most comprehensive task is the actual enrichment of shared documents with semantics. It is well known fact that people do not like to make this effort, especially not in case if somebody else has an advantage of it. As the enrichment with semantics does not solve everything in P2P networks, one has to define some basic form of an assessment of proposed solutions. Following are some issues which should be regarded in sense of effectiveness of usage of future semantically enriched P2P networks:

- **Speed (bandwidth)**: The speed of download of a document is the most important factor defining how quick the document can actually be used. This factor depends heavily on the bandwidth at the disposal of both peers, but not alone on it. Waiting in the queue for a needed file or document might take most of the time on the popular P2P networks today.
- **Effort of semantic enrichment**: The semantic enrichment of documents residing at the sites of peers raises the question of who or what is going to enrich them with semantics. With the growth of downloadable documents at peers sites the amount of the needed metadata will also grow.
- **Reusability of produced semantics**: In time ahead of the birth of the Semantic Web one should think about the reusability of semantic enriched documents therefore. There is still no certainty which technologies will at the inception of Semantic Web be active (RDF, Topic Maps, Curl), but likelihood is big that an XML-based format will be used.
- **What user group will be able to employ this model**: Due to the size of the bandwidth today at the private users with flat rates the guess is, that only institutions with leased lines, like universities or big companies, will be able to use services like exchange of huge multimedia contents. These interests shall be targeted first.
- **Billing model**: In the case of companies using these services or maybe even private universities, one should anticipate that these will not provide their contents for free. In this case introducing of a billing model should be considered outright.

The rate at which content can be transferred from a peer to another peer depends on the bottleneck bandwidth (the lowest bandwidth part of the path between the peers) between them, the available bandwidth along the path and latency between the peers [MES2002]. This is why greater upstream bandwidth of a content providing peer does not guarantee a quicker download by the content consuming peer. It only ensures that the first hop away from the data sourcing machine is not the bottleneck. However it seams reasonable that the readiness of peers to take long download periods into account would increase, if they know for sure that they are getting exactly the content that they looking for.

## 6. Conclusion

The awareness about the assertions made above raises a necessity for concepts where usage will be restricted to some special users and not for general public, like in the present P2P networks. The most adequate users of these services would be mostly educational institutions cooperating with each other on new learning contents and methodologies and to some lesser extent companies with interests and needs in IT services. At the same time one should not pursuit proprietary solutions fitting only one domain of corporate or educational needs, but aspire at a wide usage of emerging standards which will probably make the backbone of a future usage and re-usage of documents and objects.

This paper shows some important interaction of several factors in P2P networks regarding an anticipated semantic enrichment. We showed what types of institutions may have adequate resources and interests in using powerful semantic enhanced P2P networks. In anticipation of wider furnishing of learning contents with multimedia features one can expect a growing desire for proliferation of these contents among universities and similar educational bodies. We showed some implications of semantic enriched P2P networks which should be addressed outright at the their modeling. Generally there is a great potential in P2P network oriented approach considering the rising quantity of network aware computers. This technology will probably experience even wider utilisation and acceptance in the course of further decentralized management of contents on the net.

## References


DAO2001 http://www.w3.org/TR/daml+oil-reference

EDU2002 Wolfgang Nejdl, Boris Wolf, Changtao Qu, Stefan Decker, Michael Sintek, Ambjörn Naeve, Mikael Nilsson, Matthias Palmer *Edutella: A P2P Networking Infrastructure Based on RDF*, May, 2002

FRE2001 Freenet, http://freenet.sourceforge.net

GNU2000 Gnutella, http://gnutella.wego.com

JXT2001 Sun Microsystems, JXTA v2.0 Protocols Specification, http://spec.jxta.org

LOM2002 IEEE 1484.12.1 LOM draft standard, http://ltsc.ieee.org

MES2002 Stefan Saroui, P. Krishna Gummadi, Steven D. Gribble, Dept. of Computer and Engineering, Univ. of Washington, Seattle, WA, 98195-2350, *A Measurement Study of Peer-to-Peer File Sharing Systems*, November, 2002

OWL2002 http://www.w3.org/TR/owl-semantics/

RDF99 http://www.w3.org/RDF/

SET2001 SETI@home, The Search for Extraterrestrial Intelligence, http://setiathome.ssl.berkely.edu

SOV2002 Arturo Crespo and Hector Garcia-Molina, Computer Science Department, Stanford University, CA 94305-2140, USA, *Semantic Overlay Networks for P2P Systems*, 2002

Ste2002 Achim Steinacker, *Medienbausteine für web-basierte Lernsysteme,* Dissertationsschrift, Technische Universität Darmstadt, 2002

THL2001 Tim Berners-Lee, James Hendler, Ora Lassila, *Semantic Web*, Scientific American, May 2001

XIN2002 Apache Xindice, http://xml.apache.org/xindice/

XTM2001 http://www.topicmaps.org/xtm/1.0/