A New Document Sharing System Based on a Semantic Hierarchical Peer-to-Peer Network

Anis Ismail, Aziz Barbar, Mohammad Hajjar, and Mohamed Quafafou

***Abstract*—Peer-to-Peer (P2P) applications have been widely used in last years. A P2P application is usually used for sharing Music, Movies, Games, and other files. These applications work by permitting to a Peer to download files by assembling it from multiple sources on the network at the same time. In this paper, we present a new application that permits to share enriched scientific documents. We have developed a P2P application based on community architecture structured around Peers and Super-Peers. This application allows the sharing of references between researchers from different communities. The references are shared by researchers eventually augmented by annotations. Annotation allows a researcher to comment or give an opinion on a specific reference. This application was applied at the Lebanese University (IUT, Sidon) to allow instructors to share their annotations, comments, and other information.**

***Index Terms*—Peer-to-peer, scientific documents, annotations, communities.**

I. Introduction

The terms “Peer” or “Node” designated a normal computer where the term "Peer-to-Peer" (P2P) refers to the use of resources distributed over Peers connected by a network to perform tasks in a decentralized manner. The term “P2P network” means the connected Peers through ad hoc connections where all nodes have similar capabilities (Fig. 1). For this reason, each node acts as a server and a client, and Peers are often referred to as the server name.

We distinguish three types of P2P networks [1]:

- The Pure Networks: In this type of networks, all Peers in the network play a similar role. Each Peer is connected to a random subset of neighboring nodes. For example, in the system Gnutella [2], a query sent by a Peer is treated by all its neighbors and then spread throughout these connections. We talk about decentralized P2P (Fig. 1).
- The Hybrid Networks: In this type, the communication is done via a central server. For example, in the Napster system [3], a query is processed by the central node of the system to provide an answer. We are talking about centralized P2P (Fig. 2).

Manuscript received January 12, 2016; revised March 15, 2016.

Anis Ismail and Mohammad Hajjar are with the Lebanese University, University Institute of Technology, Sidon, Lebanon (e-mail: anismaiil@ul.edu.lb, m_hajjar@ul.edu.lb).

Aziz Barbar is with the American University of Science and Technology, Beirut, Lebanon (e-mail: abarbar@aust.edu.lb).

Mohamed Quafafou is with LSIS, Domaine Universitaire de Saint-Jérôme, Avenue Escadrille Normandie-Niemen, 13397 Marseille Cedex 20, France (e-mail mohamed.quafafou@univ-amu.fr).

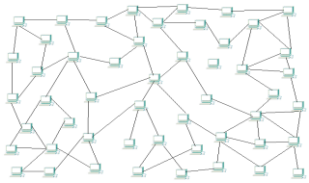

Fig. 1. Example of P2P network of 50 nodes showing the ad hoc structure.

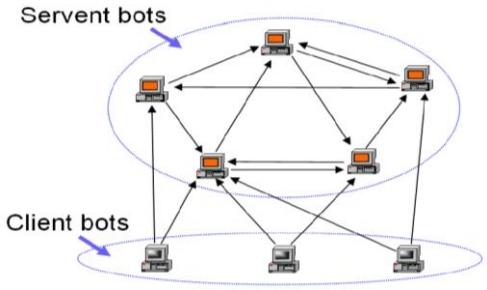

Fig. 2. Hybrid networks.

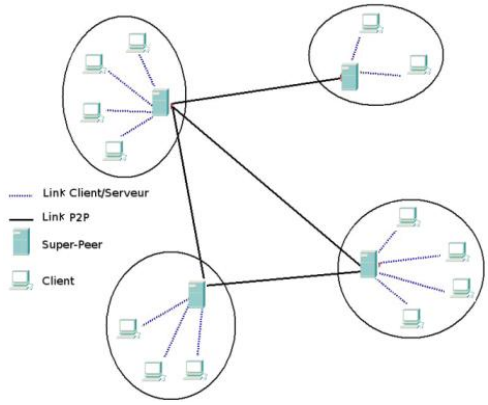

Fig. 3. Hierarchical P2P systems.

- The Hierarchical P2P Networks: In this type of network, the Nodes are classified in two categories: Peers and Super-Peers. The Super-Peers (powerful nodes) play a specific role and have processing strength and larger bandwidth than other Peers (normal nodes) of the network. The Super-Peer model [4] introduces a hierarchy between a Super-Peer and the Peers connected to this Super-Peer. The Super-Peers work in P2P mode, so that within a group, a Super-Peer and its Peers work in a classic client-server mode (Fig. 3). The hierarchical model has the advantages of using both types of systems (centralized and decentralized). A Super-Peer acts as a centralized repository for the account of a set of Peers. Routing in Super-Peers networks is more effective than pure P2P networks because the routing is limited to Super-Peer networks. This solution solves the problem of scalability of the purely distributed approach while maintaining the effectiveness of the centralized solution.

P2P systems are widely used for sharing data or documents [5]. File sharing networks like Gnutella [2] is a good example

of scalability and reliability. In Gnutella, Peers are first connected to a flat overlay network, in which every Peer is equal. Peers are connected directly without the need of a master server's arrangement and the malfunction of any node does not cause any other nodes in the system to malfunction as well.

The contribution of this paper is in developing a P2P application based on community architecture structured around Peers and Super-Peers. This solution is based on JXTA application and allows the sharing of references between researchers from different communities. The references are shared by researchers eventually augmented by annotations. Annotations allow researchers to comment or give an opinion on a specific reference. In this application, a data source is represented as an XML document (or a relational database) containing all references a researcher wishes to share with other researchers (with specific annotations). A researcher expresses XML documents using a specific format or other available formats such as JabRef, etc. A researcher belonging to a community can query the schema of the data source to retrieve information from other communities.

The following section presents the related work. Section III recalls briefly main concepts of P2P networks and shows the context of our work. Section IV, presents the implementation of our application. In Section V, we conclude the work.

II. Related Work

P2P applications are quickly emerging as large-scale systems for information sharing through networks. These applications can be classified in the following 3 categories: the Semantic Overlay Networks (SON), the Peer Data Management Systems (PDMS), and the hybrid one.

SON is an overlay network, associated with a concept of a classification hierarchy. The system SQPeer [6] is a system that uses a SON formed by grouping Peers sharing similar information on their schemes. In SQPeer, each Peer has a source of data in RDF format in accordance with RDF schemas. Queries are expressed in RQL (RDF Query Language) a SQL-style language for RDF. Each Peer publishes an RVL (RDF View Language) describing the schema. These views are shared across the P2P network. Semantic topology is used to group Peers sharing similar patterns. Each query is compared with the local views held by the Peer then it is annotated with information on the location of relevant Peers. The problem of such systems can be described as follows: given a set P of Peers physically linked, containing sources A of autonomous and heterogeneous data, we want to investigate the data from those Peers as if they were one source based only on a network N of semantic mapping. A semantic mapping defines the conceptual equivalence among attributes in schemas of different Peers. There is no process by routing queries, since the Peers to which propagates a query are determined by matching schemes.

The Edutella system [7] produces a P2P infrastructure that supports RDF metadata. The ontologies that describe the data are stored in a database or an XML document. The Edutella's topology is a Super-Peer type in which the Super-Peers are organized into hyper-cube to route queries. Edutella proceeds by diffusion in Super-Peer level. The procedure of query processing authorizes queries plans containing the predicates of selection, aggregate functions and joins.

The Bibster system [1] aims to share bibliographic data among researchers. Two ontologies can structure automatically data of Peers. Ontologies involved in data storage, reformulation, routing queries, and presenting results. The Peer selection is based on relevant expertise from Peers, leading to the formation of a semantic P2P network independent of the existing P2P topology.

The PDMS is a natural convergence between P2P systems and distributed databases. Thus, PDMS can be seen as an evolution of distributed databases to a wide distribution. The system PeerDB [8] allows the sharing of distributed relational databases. The multi-agent systems are combined with P2P systems in PeerDB. Each Peer shares this data as a relational database described by keywords. To find the relevant Peers, this Peer broadcasts its query to all its neighbors that do match keywords describing the relations of the query with those describing the relationships it has. Once done, these matching relationships with keywords will be resent to the initiator Peer that will update the query accordingly and sent to relevant Peers.

The system AmbientDB [9] is based on the PDMS approach. Each Peer has its own schema and provides mappings with the global schema existing in the system. The routing of Queries is initialized by a protocol in each Peer and uses Chord to connect the Peers between them with index table distributed among Peers. A Peer expresses its application in the form of standard relational algebra. A Super-Peer approach to AmbientDB is presented in [10].

The hybrid architecture consists of a central server which keeps information about the network. The XPeer system [11] is a hybrid architecture for sharing XML data. Each Peer exports data description to share in the form of a tree. Peers are logically organized into groups (Super-Peer) based on the similarity of patterns. The queries are written as XQuery.

III. Theoretical Framework

A. Basic Notations

A Peer is an autonomous entity with a capacity of storage and data processing. In a computer network, a Peer may act as a client or as a server. A P2P is a set of autonomous and self-organized Peers (P), connected together through a computer network. The purpose of a P2P network is the sharing of resources (files, databases) distributed on Peers by avoiding the appearance of a Peer as a central server in this network [12], [13]. We note: P2P = (P, U), P is the set of Peers and U represents the links (overlay connections) between two Peers P_i and P_j, $U \subseteq PxP$. The Super-Peer based (P2Ph) (Fig. 4) network that we consider in this paper includes sets of Peers (P) and Super-Peers (SP). We note : P2Ph = ($P \cup SP$, K), where P is the set of Peers, SP is the set of Super-Peers and K is the set of overlay links expressed under the format of pairs: (P_i, SP_j) or (SP_j ,SP_k) which respectively link a Peer P_i to a Super-Peer SP_j or a Super-Peer SP_j to one or several Super-Peers SP_k.

B. Expertise, Mapping, and Domain

We only consider data models supported by Peers. We distinguish the three following data models, the best known: relational, XML, and object. An expertise is defined, in our case, as (a part of) the data schema, expressed with one of the three data models cited above, possessed and published by a Peer in order to share its data with other Peers. To facilitate the reconciliation between the data schema of the Peer and the theme described by a Super-Peer, two measures were taken: 1. the expertise of a Peer is expressed with the language of its Super-Peer (i.e. concept, role and IsA); 2. The expertise of a Peer is expressed under the format of couple of elements, satisfying the following condition:

$$EXP(P_i) = \{ \theta(si; sj) \in SP \mid (s_i; s_j) \wedge \theta \in R\} \quad (1)$$

In our context, mapping is an important process in order to share data between Peers. Two levels of mapping are distinguished: the first level is to share data between Peers, it is important to search for connections between expertise of Peers and the description of themes provided by Super-Peers. The second level is to process users' queries, we search for connections between the subject of a query (detailed below) and the expertise of each (Super-)Peers in order to know its capacity to respond to this query. Let S_1, the expertise of a Peer and S_2 the theme proposed by the Super-Peer. The search for correspondence between S_1 and S_2 is made to find for each concept or role in S_1 (or S_2) a correspondent in S_2 (or S_1) which is the nearest semantically. We can define the concept of mapping (Map) between schemas as follows:

$$Map: S1 \rightarrow S2 \; Map(es1) = es2 \quad \text{if} \quad (2)$$

$Sim(es1; es2) > \text{acceptable-threshold}$

where es_1 is the entity of schema S_1; es_2 the entity of schema S_2; and $Sim(es_1; es_2)$ is a function that measures the similarity between two entities es_1 and es_2, given as follows:

$$Sim: S1xS2 \rightarrow [0; 1] \quad (3)$$

We distinguish two particular cases: $Sim(es_1; es_2) = 1$ describes two similar entities ; $Sim(es_1; es_2) = 0$ describes two distinct entities.

We introduce the two concepts, Semantic Intra-Domain and Semantic Inter-Domain. A Semantic Intra-Domain is an interest domain in which mappings between Peers, members of this domain, and the Super-Peer responsible for this domain are established. A Semantic Inter-Domain is a set of semantic Intra-Domain in which mappings between Super-Peers of these domains are established.

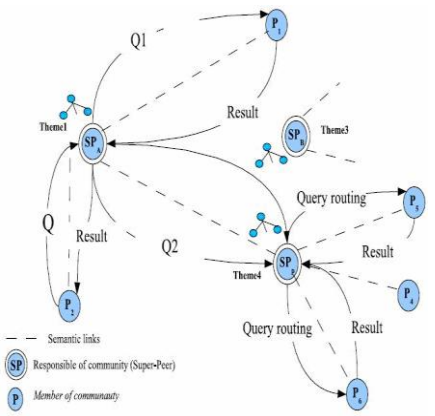

Fig. 4. Network configuration.

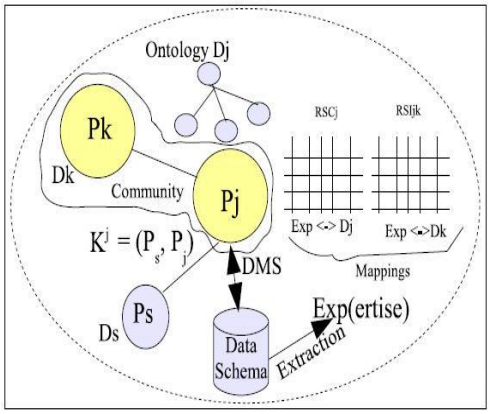

Fig. 5. Example of semantic inter-domain.

We note Semantic Intra-Domain (SI_a^jC) and Semantic Inter-Domain (SI_a^jC) number j (Fig. 5) as follows:

$$SI_a^jC = (PS \bigcup SPTj,Dj, EXP(PS), Kj; RSCj) \quad (4)$$

$$SI_a^jC = (SI_a^jC, RSIj,1, \ldots, RSIj, k), k \neq j \quad (5)$$

where $P_S \subseteq P$ is a subset of Peers having the same center of interest T_j, EXP (PS) is the set of expertise for Peers that are interested by this theme and that joined this domain. $SP_{Tj,Dj}$ (belongs to SP) is the Super-Peer responsible of the domain j which is joined by Peers (i.e. a Peer of a domain may request to join several domains if the user thinks that his/her theme of interest is in the intersection of several domains). D_j represents the description of the theme T_j provided by the Super-Peer. $K^j \subseteq K$ is the set of overlay links between the Super-Peer $SP_{Tj,Dj}$ and the Peers connected to it combined with the set of overlay links between $SP_{Tj,Dj}$ and Super-Peers $SP_{Tk,Dk}$, $k \neq j$. RSCj is the semantic Intra-Domain between the Super-Peer $SP_{Tj,Dj}$ and the Peers inside this Domain. $RSI^{j,k}$ is the semantic Inter-Domain concerning the links found between the description of the theme D_j of the Super-Peer $SP_{Tj,Dj}$, with the description D_k of each Super-Peer $SP_{Tk,Dk}$, $k \neq j$). Finally, we introduce a SON represented by the union of all the semantic networks of intra-Domains and inter-domains. A SON is noted as follows:

$$SON = \bigcup_{j=1}^{|T|} (SI_e^jC) \quad (6)$$

where *T* represents the total number of Super-Peers in the network. The next section will present the query routing algorithm (our baseline approach).

IV. Implementation of the Application

This application has been tested at the Lebanese University (IUT, Sidon). The IUT consists of 4 academic departments (IAG, GRIT, GIM, GC) with an administration department (DIR).

Each researcher/instructor of IUT is attached to a department. In our context, each teacher designates a Peer and each department is a Super-Peer (Fig. 6). Each Peer is attached to the Super-Peer pointing its department.

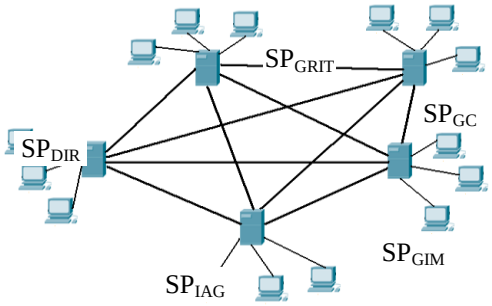

Fig. 6. P2P network of 5 SP.

Peers of application share course materials, books, reports, scientific publications (references) as well as administrative documents. Indeed, there are instructional materials that are shared by all departments (human rights, expression and communication, ...) or that are common between four departments. For example, materials like RDB (Relational DataBase), OOP (Object-Oriented Programming), AGPG (algorithms and programming), SE (Software Engineering), Web Statistics, Probability, Networks, ... are shared between the both IAG and GRIT departments.

This application allows a user to provide others with publications that he/she has on his/her local machine; these publications can be indexed in different ways (keywords, authors, title, abstract, etc.). Symmetrically, the system must allow a user to perform queries to find a specific resource (using the same criteria indexing). Each Peer acts as a client when it sends a query, and as a server when it responds to a query from other users.

We consider two Super-Peers SP1 and SP2. The two Super-Peers have each a schema (ontology) represented in the sGraph format. One of these Schemas is given in Fig. 7.

- The schemas of the Super-Peer SP1 describes the publications stating the author, title, keywords, and type of publication (conference, book, or journal), annotations etc..
- The schemas of the Super-Peer SP2 describes only the publications in conferences indicating the name of the conference, the country where it took place, the date and the title and author of each publication etc..

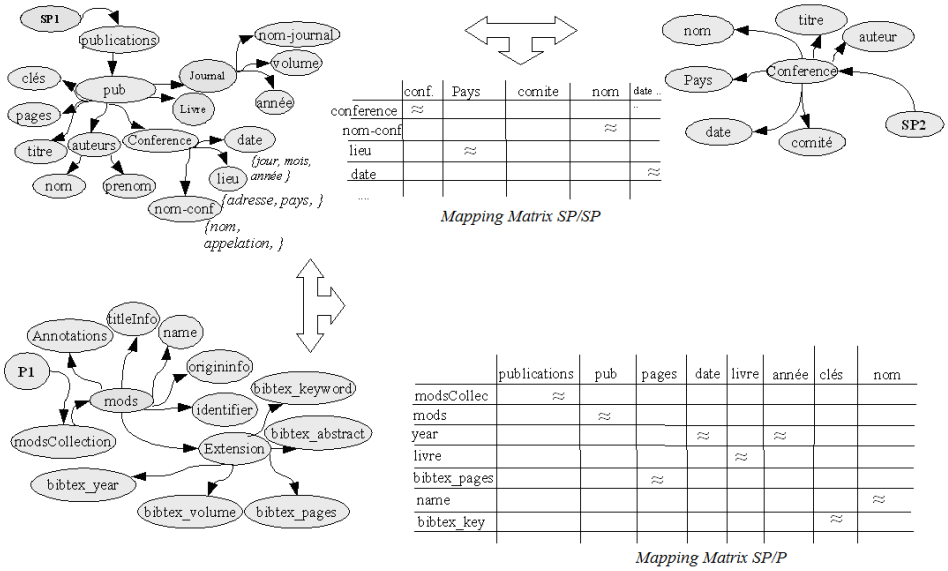

Fig. 7. Schemas.

```
<?xml version="1.0" encoding="utf-8" ?>
<!DOCTYPE modsCollection SYSTEM "test2.dtd">
<modsCollection>
<mods ID="ElAssad2006" version="3.0">
<titleInfo>
<title>WBLST - E-Learning </title>
</titleInfo>
<name type="personal">
<namePart type="family">Assad</namePart>
<namePart type="given">S. El</namePart>
<role>
<roleTerm type="text">author</roleTerm>
</role>
</name>
<originInfo>
<dateIssued>2006</dateIssued>
<issuance>monographic</issuance>
</originInfo>
<identifier>ElAssad2006</identifier>
<typeOfResource>text</typeOfResource>
<genre authority="marc">Inproceedings</genre>
<relatedItem type="host" version="3.0">
<titleInfo>
<title>Proc. 2nd Information and</title>
</titleInfo>
<originInfo>
<issuance>continuing</issuance>
</originInfo>
<typeOfResource>text</typeOfResource>
<extent unit="page">
<start>690</start>
<end>694</end>
</extent>
</relatedItem>
<extension>
<bibtex_owner>olivier</bibtex_owner>
</extension>
<extension>
<bibtex_timestamp>2008.12.14</bibtex_timestamp>
</extension>
<extension>
<bibtex_year>2006</bibtex_year>
</extension>
...
<annotations>Permet d'avoir plusieurs méthodes avec même nom
</annotations>
</mods>
</modsCollection>
```

(a) XML document

```
<!ELEMENT modsCollection ( mods+ ) >
<!ELEMENT mods ( titleInfo, name+, originInfo, identifier,
          typeOfResource, genre, relatedItem, extension+, annotations>
<!ELEMENT bibtex_abstract ( #PCDATA ) >
<!ELEMENT bibtex_booktitle ( #PCDATA ) >
<!ELEMENT bibtex_doi ( #PCDATA ) >
<!ELEMENT bibtex_file ( #PCDATA ) >
<!ELEMENT bibtex_journal ( #PCDATA ) >
<!ELEMENT bibtex_keywords ( #PCDATA ) >
<!ELEMENT bibtex_owner ( #PCDATA ) >
<!ELEMENT bibtex_pages ( #PCDATA ) >
<!ELEMENT bibtex_review ( #PCDATA ) >
<!ELEMENT bibtex_timestamp ( #PCDATA ) >
<!ELEMENT bibtex_volume ( #PCDATA ) >
<!ELEMENT bibtex_year ( #PCDATA ) >
<!ELEMENT dateIssued ( #PCDATA ) >
<!ELEMENT end ( #PCDATA ) >
<!ELEMENT extension ( bibtex_abstract | bibtex_booktitle |
          bibtex_doi | bibtex_file | bibtex_journal | bibtex_keywords
          | bibtex_owner | bibtex_pages | bibtex_review
          | bibtex_timestamp | bibtex_volume | bibtex_year )* >
<!ELEMENT extent ( start, end ) >
<!ATTLIST extent unit NMTOKEN #FIXED "page" >
<!ELEMENT genre ( #PCDATA ) >
<!ATTLIST genre authority NMTOKEN #FIXED "marc" >
<!ELEMENT identifier ( #PCDATA ) >
<!ELEMENT issuance ( #PCDATA ) >
<!ATTLIST mods ID NMTOKEN #REQUIRED >
<!ATTLIST mods version NMTOKEN #FIXED "3.0" >
<!ELEMENT name ( namePart+, role ) >
<!ATTLIST name type NMTOKEN #FIXED "personal" >
<!ELEMENT namePart ( #PCDATA ) >
<!ATTLIST namePart type ( family | given ) #REQUIRED >
<!ELEMENT originInfo ( dateIssued | issuance )* >
<!ELEMENT relatedItem ( extent | originInfo | titleInfo
          | typeOfResource )* >
<!ATTLIST relatedItem type NMTOKEN #FIXED "host" >
<!ATTLIST relatedItem version NMTOKEN #FIXED "3.0" >
<!ELEMENT role ( roleTerm ) >
<!ELEMENT roleTerm ( #PCDATA ) >
<!ATTLIST roleTerm type NMTOKEN #FIXED "text" >
<!ELEMENT start ( #PCDATA ) >
<!ELEMENT title ( #PCDATA ) >
<!ELEMENT titleInfo ( title ) >
<!ELEMENT typeOfResource ( #PCDATA )>
<!ELEMENT annotations ( #PCDATA ) >
```

(b) DTD

Fig. 8. DTD for peer P1.

The semantic reconciliation between Schemas of Super-Peers is given in the correspondence matrix SP/SP. We assume that each node (in sGraph) is associated with a set of synonyms. This set helps the search of semantic links (semantic reconciliation) between schemas (sGraph) Super-Peers.

We also consider the schema of a Peer P1 expressed with sGraph. This schema describes references in the JabRef format. JabRef is a software of reference management that uses BibTeX as native format. JabRef provides an easy editing BibTeX files. It allows you to import references from the Web.

Reconciliation between the schema of Peer P1 and schema pattern of its godfather (the Super-Peer SP1) is also established. The result part of this reconciliation is given in the correspondence matrix SP/P.

In Fig. 8, we present the data source of Peer P1. This source is expressed with the JabRef format. It contains annotated references that researcher wishes to share with other researchers through the P2P network. Fig. 8 shows the Document Data Definition (DTD) of the source.

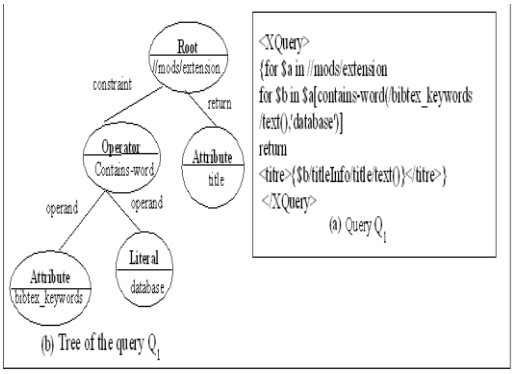

Fig. 9. Query Q1.

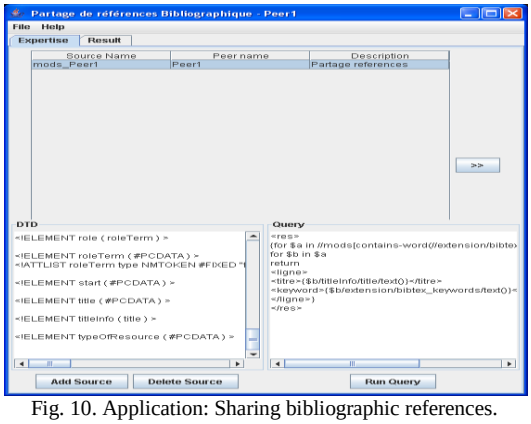

Fig. 10. Application: Sharing bibliographic references.

The user of Peer P_1 expresses queries using the schemas of his/her DTD. In our case, the DTD is retrieved automatically by the Peer from a data source. Query Q_1 shown in Fig. 9 page 5, selects the network references that contain the keyword 'database'.

Q_1:

<Xquery>

{for $a in //mods/extension

for $b in $a[contains-word(/bibtex_keywords/text(),'database')]

return

<titre>{$b/titleInfo/title/text()}</titre>}

</XQuery>

Q1 is represented in a tree, its subject is given as follows:

Sub(Q_1)= {(//mods/extension,bibtex_keywords),(//mods/extension,title)}

This query is then sent to Super-Peer SP_1 seeking Peers and

Super-Peers that are able to process it. One of the mentioned three approaches can be used to obtain the final result and return it to the user as indicated in Fig. 10.

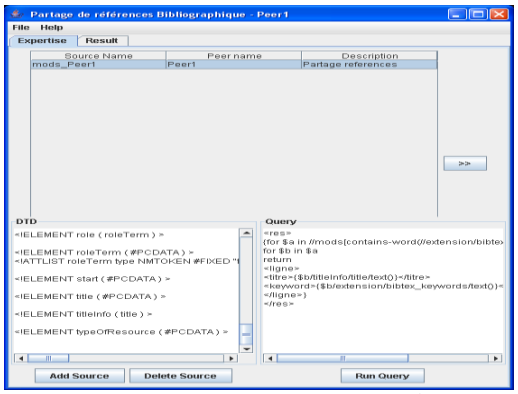

Fig. 10. Application: Sharing bibliographic references.

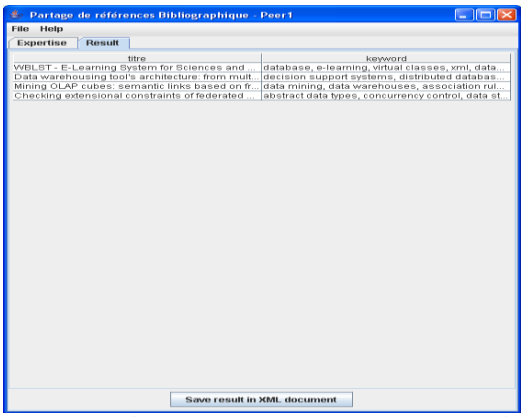

Fig. 11. Results of query Q1 returned by the application interface.

The following Query Q_2 consists of selecting from the community network the annotations concerning the reference "SON".

Q_2 :

<Xquery>

{for $a in //mods/titleInfo

for $b in $a where $a/title/text() = 'Semantic Overlay Network'

return

<annotations>{$b/annotations/text()}</annotations>}

</XQuery>

The interface that we developed is shown in Fig. 8. This interface supports various functions such as the expression of queries using XQuery and visualization of the returned results.

In Fig. 11 page 6, we find the name of the data source (mods) to publish on the network to share with other Peers, the name of the Peer that publishes this data source (Peer1), and a brief description of the source. The user of Peer1 loads the data source to share (the XML document in Fig. 8) through the interface (Add source button).

The interface of Peer1 extracts automatically the schema of the source of XML data in an sGraph that is presented to

the user in a DTD format (lower left side of the interface in Fig. 10 page 6). The arrow allows sending the schema of Peer to its Super-Peer. Then the user can formulate its query based on its DTD data source.

V. Conclusion

P2P networks open a new channel for efficient downloading and sharing of files and data. P2P applications have evolved from simple, centralized, music sharing services to complex, decentralized, and file exchange mechanisms. The current crop of P2P applications are capable of exposing corporate information, damaging data, consuming resources, and stealthily tunneling straight through the firewall and proxy server. The most important key in scientific research is the sharing of information between researchers in different communities. This sharing is usually done through scientific publications (journal articles, conference papers, books, etc). In this paper, we implemented a P2P system that allows multiple computers to communicate over a network and share objects - most often files. Future work will include continuous multimedia streams (streaming), distributed processing, telephony (like Skype), etc. over the internet.

Acknowledgment

This work has been done as part of the project "Extraction des Communautés dans les réseaux sémantiques pairs-à-pairs: Simulation et évaluation" at the Lebanese University.

References

[1] J. Broekstra, M. Ehrig, P. Haase, F. Harmelen, M. Menken, P. Mika, B. Schnizler, and R. Siebes, "Bibster: A semantics-based bibliographic Peer-to-Peer system," in *Proc. the 3rd International. Semantic Web Conference*, Hiroshima, Japan, pp. 122–136, 2004.

[2] Gnutella protocol. (2000). [Online]. Available: http://rfc-gnutella.sourceforge.net/developer/ testing/index.html

[3] Napster. (2000). [Online]. Available: http://www.napster.com/

[4] B. Yang, "Garcia-Molina, designing a super-peer network," in *Proc. the 19th International Conference on Data Engineering (ICDE'03)*, IEEE Computer Society, Bangalore, India, pp. 49–60, 2003.

[5] D. Faye, G. Nachouki, and P. Valduriez, "Semantic query routing in senPeer, a p2p data management system," in *Proc. 1st International Conference on Network-Based Information Systems*, Regensburg, Germany, Sept. 3-7, 2007.

[6] G. Kokkinidis and V. Christophides, "Semantic query routing and processing in p2p database systems: The ics-forth sqPeer middleware," *EDBT Workshops*, Heraklion, Crete, Greece, pp. 486–495, 2004.

[7] W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmér, and T. Risch, "Edutella: A p2p networking infrastructure based on rdf," in *Proc. the 11th International Conference on World Wide Web*, New York, NY, USA, pp. 604–615, 2002.

[8] W. Iong, B. C. Ooi, K. L. Tan, and A. Zhou, "Peerdb: A p2p-based system for distributed data sharing," in *Proc. the 19th International Conference on Data Engineering (ICDE'03)*, pp. 633–644, 2003.

[9] A. Peter, Boncz, and C. Treijtel, "Ambientdb: Relational query processing in a p2p network," in *Proc. the International Workshop on Databases, Information Systems and Peer-to-Peer Computing (DBISP2P)*, Springer Verlag, vol. 2944, pp. 153–168, 2003.

[10] S. Brahmananda, "Design of peer-to-peer protocol for ambientdb," Master's thesis, University of Twente, 2003.

[11] C. Sartiani, P. Manghi, G. Ghelli, and G. Conforti, "XPeer: A selforganizing xml p2p database system," in *Proc. the EDBT 2004 Workshop on Peer-to-Peer Computing and Databases*, Heraklion, Crete, Greece, pp. 456–465, 2004.

[12] A. Ismail, M. Quafafou, G. Nachouki, and M. Hajjar, "Data mining effect in peer-to-peer queries routing," in *Proc. the International Conference on Management of Emergent Digital Eco-systems*, pp. 65-72, 2009.

[13] A. Ismail, M. Quafafou, G. Nachouki, and M. Hajjar, "Efficient super-peer-based queries routing," in *Proc. the International Conference on Management of Emergent Digital Eco-systems*, pp. 91-98, 2009.

Anis Ismail was born in Lebanon in April 1979. He is an associate professor of computer science at the Lebanese University, University Institute of Technology, Sidon, Lebanon. He received his Ph.D. in computer science in 2010 from Aix-Marseille University. He has a B.S. degree in telecommunication and networking engineering from the Lebanese University (LU), an M.S. in computer science from the American University of Science and Technology (AUST), Lebanon. His research interests include data/knowledge based systems including machine learning, data mining, web services, data management in peer-to-peer systems, and mobile applications.

Aziz Barbar is the dean of the Faculty of Arts and Sciences Science at the American University of Science & Technology (AUST), Lebanon. He has a Ph.D. in computer science from the University of Nice-Sophia Antipolis, France. His research interests include database reverse engineering, data mining and natural language processing. Dr. Barbar is currently the vice-president of the Lebanese Information Technology Association (LITA), and the vice-chair of the IEEE section, Lebanon.

Mohammad Hajjar is a professor at University Institute of Technology, Lebanese University, in Lebanon. He received a PHD in computer Science at Nantes University in France. His interest domain concerns Arabic language processing, multimedia information research and data management in peer-to-peer systems.

Mohamed Quafafou did his PhD thesis in 1992 on intelligent tutoring systems at INSA de Lyon, France. From 1992 to 1994, he was ATER at INSA de Lyon and than at Nantes Faculty of Sciences. From 1995 to 2001, he was assistant professor at the Nantes University. During that period, he developed research on rough set theory, concepts approximation, data mining, web information extraction and participated actively with France Telecom to the project comminges to design a new web system dedicated to French web analysis for discovering emergent web communities. He was also chief-scientist at GEOBS where he headed the Geobs Data Analyzer project, which was developing a spatial data mining systems with application to environment, marketing, social analysis, etc. From September 2002, he was professor at the Avignon University and moved in 2005 to the Aix-Marseille University where he joined the Information and System Science Laboratory (UMR CNRS 6168) and continue his research on web data mining considering different application contexts like P2P, multimedia and web services. Since 2002, he teaches foundations of data/knowledge based systems including machine learning, data mining, personalization, data warehousing, XML, web services, multimedia, web and mobile applications.