Towards Peer-to-Peer Semantic Web: A Distributed Environment for Sharing Semantic Knowledge on the Web

Madhan Arumugam, Amit Sheth, and I. Budak Arpinar
Large Scale Distributed Information Systems Lab.
Computer Science Department, University of Georgia
http://lsdis.cs.uga.edu, <arumugam, amit, budak>@cs.uga.edu

Abstract. The real value of Semantic Web vision can be demonstrated if people and applications can create and discover new and interesting knowledge and share this knowledge in a transparent manner similar to the way data is exchanged today. Therefore, we believe that Semantic Web will consist of a distributed environment of shared and interoperable ontologies, which have emerged as common formalisms for knowledge representation. The users will need to discover new ontologies, which are not known to them before and use them to either annotate the content or to formulate their information requests. This requires an environment that supports creating, maintaining, and controlled sharing of ontologies. We believe that a Peer-to-Peer (P2P) infrastructure can enable such capabilities. We call our approach P2P Semantic Web (PSW) with capabilities to find relevant set of ontologies, facilitating reuse of existing ontologies to create additional ontologies, and advertising the resulting ontologies. We discuss an ontology driven search of concepts and services and exploration of inter-ontological relationships over a P2P infrastructure. The prototyping of this approach in the InfoQuilt system is also discussed.

1. Introduction
Semantic Web has been presented as the next step in the evolution of the World Wide Web [B99, BHL01, FM01]. Central to the current Semantic Web vision is the promise that the users will be able to use programs that can understand semantics of the data. In support of such a vision, a good amount of recent work has focused on annotating data using ontologies to make data machine-understandable. From a broader perspective, Semantic Web is intended to make the Web more intelligent by improving the value of, and interactivity with vast amount of data on the Web. One approach for such a support for semantics is that of using ontologies to provide the context for capturing the meaning of data as well as the user’s intention in a query. Such an approach has been advocated for some time (e.g., [KS94, MKSI96]). Eventually, this can help users answer complex questions like find a correlation between earthquakes and nuclear tests, from data of different types, formats, and media from multiple resources.

There are many challenges we need to address in order to realize the above vision. We see a need for an environment where people and applications can share knowledge, locate new knowledge and use this knowledge in a transparent manner similar to the way data is exchanged today. Difficulty in creating and maintaining large ontologies and the need for knowledge sharing and ontology interoperation has been presented in [MWD01, KS94]. One challenge is that of advertising knowledge and ontologies for different information domains, which are maintained by different persons, groups and organizations on the Web. For example one needs to easily locate ontologies for earthquake and nuclear test domains in the above example. With a wide variety of communities and user groups on the web, it is not trivial to find most relevant set of ontologies. A semantic search mechanism is needed to find most relevant set ontologies using users’ context for information request and his profile. Once the ontologies are located there is a need for introducing some relationships across ontologies and supporting techniques for ontology interoperation [MWD01, MIKS00]. Finally, users need tools that would allow them to define information requests (e.g., find a site for a new landfill near the main source of the municipal refuse, where the earthquake’s impact must also be evaluated) or knowledge discovery involving hypothesis validation (e.g., do nuclear tests effect earthquakes? If so find the correlation), and to collaborate with other users involved in similar knowledge-creation activities.

The InfoQuilt System developed at the LSDIS Lab. provides a framework for formulating complex information requests, which can capture the semantics of user’s request involving multiple ontologies [LSA+01], and support a form of knowledge discovery [TSP01]. Our current work involves integrating
InfoQuilt’s semantic capabilities with technologies more appropriate for developing a distributed and collaborative Semantic Web platform. As a step towards this objective, we investigate the use of Peer-to-Peer (P2P) computing as a possible infrastructure paradigm. We call the new concept a Peer-to-Peer Semantic Web (PSW). In realizing PSW, DAML+OIL provides a specification framework for independently creating, maintaining, and interoperating ontologies while preserving their semantics, and P2P is used to provide a distributed architecture which can support sharing of independently created and maintained ontologies. Our PSW facilitates:

- Distributed and autonomous creation and maintenance of local ontologies,
- Advertisement (i.e., registry) and search of (local) ontologies,
- Introducing inter-ontological relationships between relevant ontologies as-needed basis once they are located,
- Controlled sharing of knowledge base components among users in the network,
- Ontology-driven semantic search of concepts and services,
- Knowledge discovery and exploration of inter-ontological relationships.

The two important problems the PSW approach intends to address are 1) what knowledge to share, in other words semantic search of relevant ontologies and 2) how to share independently created and maintained ontologies.

The rest of the paper is organized as follows: Section 2 briefly summarizes the related work. Section 3 briefly discusses the background of InfoQuilt system and its semantic capabilities as essential components of Semantic Web. Section 4 details the semantic search of relevant ontologies with examples. Section 5 presents our conclusions.

2. Related Work

The related work on knowledge sharing primarily focuses on development of a multi-ontology environment and either integration or interoperation of these ontologies. As discussed in the literature, since ontology interoperation supports development and maintenance of locally developed independent ontologies it is superior to the integration approach. Our PSW approach is also based on the principle of multi-ontology interoperation but also facilitates an efficient search technique for new ontology discovery. In this way a proper set of ontologies are automatically located and relevant inter-ontological relationships are introduced, and this is an advancement which does not exist in the previous multi-ontology integration or interoperation approaches.

Some of the past research on knowledge sharing in a multi-ontology environment motivates our PSW approach. For example, [KS94] talked about supporting information resource discovery in a system with multiple ontologies, and [MKSI96, MIKS00] discussed processing of information requests, again involving multiple ontologies. [EFG+99, PB99] talked about approaches in using ontologies for knowledge reusing and sharing. Several issues of knowledge sharing, especially those related to ontology interoperation are discussed in [MWD01]. [STP01] presented the concept of information requests involving complex relationships and multiple ontologies, and use of parameterized information requests for knowledge discovery through what-if analysis. [GHI+01] focused on the data placement problem of how to distribute data and work among peers cost effectively. The above efforts primarily addressed the infrastructure and syntactical issues of knowledge sharing. [DHM+01] detailed about accessing information and services in a DAML enabled web, although it lacked an infrastructure to support knowledge sharing. [ERN01] proposed using P2P to support the semantic web to link semantic definitions, but lacked support for complex relationships and knowledge discovery. Our PSW approach addresses more logical and semantic level of knowledge sharing with support for exploration of inter-ontological relationships and hence knowledge discovery.
3. Knowledge Discovery in InfoQuilt

InfoQuilt is an agent-based system that allows users to semantically request information, semantically correlate data from different sources and of heterogeneous type or representation, and analyze data available from diverse autonomous and heterogeneous sources [STP01]. Furthermore, it also supports knowledge discovery through interactive what-if analysis or information hypotheses including conjectures and relationships within and across domains. InfoQuilt system includes: (a) language and tools to specify IScapes (i.e., semantic information requests), and (b) tools and algorithms to perform what-if analyses to search the information space of semantically related data. IScapes allow parameterized specification of information requests and correlation that utilizes the domain ontologies, inter-ontological relationships and user defined functions to accurately describe a user’s information need. IScapes are more than a traditional query in that they can understand user’s request by embedding semantic information [LSA+01]. For example consider the following information request:

“Find all earthquakes with epicenter in a 5000 miles radius of the location at latitude 60.790 North and longitude 97.570 East and find all tsunamis that they might have caused.”

In addition to the known explicit constraints, the system needs to comprehend what the user means by saying *find all tsunamis that might have been caused due to the earthquake*. An IScape is specified in terms of the components of the knowledge base of the system such as ontologies, inter-ontological relationships (or articulations), and user-defined operations. Representing that *an earthquake caused a tsunami* requires a complex inter-ontological relationship, involving ontologies that define earthquake and tsunami using different ontologies, and inter-ontological relationships with additional parameters and functions for spatial and temporal proximity between a pair of earthquake and tsunami events. Thus we define IScape as a computing paradigm that allows users to query and analyze the data available from diverse autonomous sources, gain better understanding of the domains and their interactions as well as discover and study relationships [STP01]. Building upon previous work, the current work changes the representation to DAML+OIL and DAML-S (for service discovery), which provides more features than our earlier XML and RDF based specification. More importantly, earlier system could execute an IScape on distributed resources, but the IScape construction was done using tools that used a single knowledge base. So a key objective of our current work is to support distributed knowledge construction and sharing of a distributed knowledge base.

4. Semantic Knowledge Sharing in InfoQuilt

4.1. InfoQuilt’s Peer-to-Peer Semantic Network

Peer-to-peer computing is the sharing of computer resources and services by direct exchange between the systems. These resources and services include the exchange of information, processing cycles, cache storage, and disk storage for files [Pee]. The popular P2P systems like Napster [Nap] and Morpheus [Mus] have attracted enormous attention towards Peer-to-Peer computing. The features in P2P that makes it desirable to be an infrastructure for knowledge sharing in Semantic Web include the following:

- It encourages distributed architecture, and supports decentralization of control and access to information and services. In this way, it provides a way to harness the computing power and knowledge of millions of computers in the web.
- It provides access to semantic information published by several independent content providers, and enables creation of personalized semantic relationships.
- It supports for publishing peer definitions and relationships to other peers and software agents.
- It offers user-centered, data-centered and computing centered models [GAR01], which provide suitable architectures for distributed content management.

The InfoQuilt System modified as part of the current work has users connected in a P2P network, accessing a shared directory of information and services. The P2P network in InfoQuilt is user centered [GAR01], where a client registers itself to a directory or directories, and searches the directory for peers providing *semantically relevant* information and services. The directory also provides the necessary contact information to connect to a specific peer, so that a direct peer-to-peer connection can be
established (see Figure 1). Currently we are using a proprietary directory service but are also investigating UDDI [UDD00] for a possible use for advertising and registering different ontologies. Each user in the P2P network runs a multi-agent Information Brokering System whose architecture is built upon the work done in [LSA+01]. The information and services that a user is interested in are semantically identified using the Knowledge Space Navigation algorithm, which provides means for locating relevant ontologies, hence relevant content. Users can build IScapes using the IScape Builder utilizing the most relevant knowledge about concepts, which are semantically identified and retrieved from the P2P Knowledge Sharing network.

An earlier work focused on supporting complex inter-ontological relationships and knowledge discovery in InfoQuilt. This work focuses on knowledge sharing in a P2P driven Semantic Web using DAML+OIL ontologies. The ontologies define concepts, its properties and its relationship with other concepts. Thus with sharing ontologies, we also share inter-ontological relationships. PSW offers the users more semantic and personalized choices in formulating IScapes, in terms of ontologies created by other peers in the network. To realize the above feature, the IScape Builder of a user now has the capability to access and add new ontologies and user-defined inter/intra-ontological relationships to the knowledge space in the P2P network. Therefore, a global knowledge space within the scope of this paper is assumed to be composed of local ontologies, which are created and maintained by independent peers and connected by some inter-ontological relationships. The user searches for a relevant ontology and the results are re-ranked based on the user profiles and query relationships by the personalization agent. The architecture of an InfoQuilt peer is shown in Figure 1.b.

4.2. Knowledge Space Construction and Navigation in InfoQuilt

The key issue while implementing the Semantic Knowledge Sharing is to construct a globally accessible knowledge space. Users must be able to do more than mere access of knowledge space. Each user must be able to add or remove concepts, without having to re-engineer the knowledge space [MUS92]. In a PSW, users can create their own ontologies subscribing to the definitions they refer to. So the users can construct their own ontologies and at the same time conform to the Global Knowledge Space. All ontologies defined in DAML+OIL are qualified by namespaces. Conceptually namespaces are identifiers or references in the knowledge space. When a person defines his own concept or notion based on a pre-defined or agreed upon concepts, it is marked up in the knowledge space by these references. It does not necessarily mean that we have to agree to all the definitions existing in the knowledge space, but DAML+OIL helps to create a democratic and ecological knowledge space where different concepts and definitions can co-exist. One that survives is the one that is most referenced, other definitions go by unnoticed. In using DAML+OIL, the only concepts everybody agrees upon are the basic classes like Thing [Thi]. These are the predefined most or least general classes of all the classes that are defined in DAML+OIL [Dam]. With the use of DAML+OIL the knowledge space is constructed from roots. When a
new ontology is created, it is already hooked up in the knowledge space because of the use of imports and namespaces [Dam]. PSW’s support for peer registration of ontologies is discussed later in this section.

In order for the programs to access this knowledge space programmatically, a data structure is used as described. A DAML+OIL ontology is a collection of RDF triples [Oil]. Each RDF statement is a resource reifying a triple. Such a resource must have at least three properties rdf:subject, rdf:object and rdf:predicate, valued by the corresponding resources. Each RDF statement in an ontology can define a class, its properties and its relationships with other classes. For example a statement would look like 

\[<\#boy> <\#drinks> <\#coffee>\].

Now all of the triples, boy, drinks and coffee are qualified by use of URIs. The data structure stores subject (boy), the object (coffee) and the verb (drinks) that relates them. So the core components of the data structure are

- Concepts: Concepts or KObjects as we refer to contains information about each class. In the example, boy and coffee qualify to be KObjects.
- Links: Links or relationships contain information about the predicate and the KObjects it relates. In our example drinks qualifies as a Link.

For each KObject, the pointers to the Links that has this KObject as a subject or object and the ontology it is defined along with the user information are maintained. In creation of the knowledge space the following steps are involved:

1. Retrieve every RDF triple (subject, predicate, and object) from each source ontology,
2. For every assertion of a fact or a definition made in the ontology, recursively trace its link to the most general class of the knowledge space (#Thing),
3. Repeat 1 and 2 untill all the ontologies are hooked into the knowledge space.

For knowledge space navigation, we can start with the KObject Thing and then traverse through the Links in the KObject.

4.3. Ontology Registration

In PSW, ontologies (DAML+OIL) are defined as a collection of definitions and assertions. The peers can create and maintain their own ontologies. They have the control as in whether or not to share an ontology. Any peer can create his or her own definitions conforming to DAML+OIL formalisms. Additionally, peers can define ontologies through assertions that refer to definitions created by other peers and also can let other peers refer to their definitions. When a peer decides to share an ontology in the global knowledge space, s/he has to upload the ontology into the PSW. This process is called registration of ontologies. New concepts (KObjects) and relationships (Links) in the uploaded ontology are created appropriately. Once ontology is registered, other peers can refer to the definitions in this ontology for their use. This approach enables controlled sharing of ontologies. For example, a peer can protect his definition by not sharing, but use assertions that refer to shared definitions. In case a peer decides to remove or deregister his ontology, all the definitions and the assertions that refer to these definitions become invalid in the knowledge space. Tools like DAML Validator can be used to check for obsolete definitions and stale links.

5. Semantic Search

One of the key advantages of constructing a knowledge space is semantic search. In the IScape Builder, the user specifies the keywords (usually common nouns) used in the information request. The next step in the construction of the IScape is selection of ontologies that help the system understand the user’s query [STP01]. In the previous version of InfoQuilt, on which our PSW approach is built, the users need to create either their own ontologies from scratch or manually select some of the previously created local ontologies. However, the PSW approach enables InfoQuilt System to search the knowledge space for the relevant ontologies.

The data structure representing the knowledge space is a collection of KObjects and Links. The process of searching involves the following steps. The input is a set of keywords and the output is a list of ontologies.
1. Take each keyword and run a basic keyword match on the subject, object and the predicate (in that particular order) in the entire knowledge space,
2. Retrieve the name of the ontologies that satisfy the above match along with the ownership details,
3. If the keywords result in a number of ontologies, compare the ontologies for common parents and eliminate the ontologies without any common links,
4. If there is more than one ontology describing the same keyword, perform search with more keywords or compare the resulting ontologies to help user select the ontology.

Even though this search is invoked using keywords, we call it semantic, because the keyword matches are done with the KObjects qualified in the knowledge space, which by very construction preserves the semantics of the knowledge it describes. Unlike ordinary keyword searches, this kind of search has specific direction, because we believe that the keywords used in a valid information request are semantically woven in the holistic meaning they represent. One other utility we have developed using the knowledge space is the ability to compare two ontologies. This involves the following steps:

1. Identify the KObjects (concepts) used in each ontology in the knowledge space,
2. Find a common parent KObject that links two KObjects that are defined in each of the compared ontologies; in other words, find a connecting link (relationship) between the two ontologies and trace it for the user.

With DAML+OIL used in defining the ontologies, any two concepts will be linked together by the fact that both of them are Things. The scope of defining how close a common KObject can be in terms of the number of Links to be defined as a certain relationship is an interesting research problem and is beyond the scope of this paper. This utility can thus help the user explore inter-ontological relationships, and hence knowledge discovery. Now we will demonstrate the search with an example.

**Example 1:** Consider the information request

“Find all earthquakes with epicenter in a 5000 mile radius of the location at latitude 60.790 North and longitude 97.570 East and find all tsunamis that they might have caused.”

The keywords in the above information request are earthquake, epicenter, radius, location, latitude, longitude, and tsunamis. Just viewing these keywords, any reader might guess the verbs and the semantics that connect the meanings of these keywords. The challenge is to make the system understand with the use of knowledge space. Let us assume the following results for the keyword matches on the subject, object and the predicate of all the triples in the DAML+OIL ontologies:

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Ontologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earthquake</td>
<td>earthquake.daml, damage.daml</td>
</tr>
<tr>
<td>Latitude, Longitude, Location</td>
<td>location.daml, weather.daml, earthquake.daml</td>
</tr>
<tr>
<td>Epicenter, Radius</td>
<td>earthquake.daml, circle.daml</td>
</tr>
<tr>
<td>Tsunami</td>
<td>tsunami.daml</td>
</tr>
</tbody>
</table>

Figure 2 shows simplified versions of ontologies, which are used in this example. After the above results are obtained, we have to arrive at the semantically relevant set of ontologies. This is done by the comparing every KObject in the ontology with every other KObject in the other ontologies. If they have a common KObject linked by both the KObjects, they are related (e.g., tsunamis and earthquakes are related because they have a common parent, i.e., a KObject, namely disaster). In our example we have six ontologies matching seven keywords. Of these ontologies earthquake.daml, location.daml and tsunami.daml are linked with KObjects latitude, longitude. So the relevant set of ontologies will be earthquake.daml, location.daml, tsunami.daml, discarding damage.daml, weather.daml and circle.daml. The ontology weather.daml is discarded even though it has a reference to the definition of location. Because, although the definition of weather involves country, which is a sub-class of location it is not related in the sense it does not have a common parent with earthquake and tsunami. Thus, the system
considers earthquake, tsunami, and location ontologies as relevant because this is the minimal set of ontologies with all keyword matches and has at least one common KObject linked.

This example also demonstrates the compare function, which compares two ontologies to trace the relation between them. For instance, consider ontologies earthquake.daml and tsunami.daml. Their links in knowledge space are latitude, longitude and disaster. This function can be effectively used for exploring inter-ontological relationships and to discover knowledge on that basis. The diagram below illustrates the ontologies and their relationships and how they are linked up in the knowledge space. This example is also illustrative of how the system can navigate between concepts in the knowledge space.

After the relevant set of ontologies are obtained from the P2P network, the personalization techniques re-rank the results according the user’s profile and query history. The user then continues with the selection of ontologies and other tasks in the IScape Builder. The IScape Builder is a stand-alone Java application that provides a graphical interface to create and execute ISscapes. This builder enables users to provide the keywords for searching ontologies. Then the user can select the ontologies s/he wants to use in the IScape. The next step for constructing the ISscapes is to define user-defined relationships based on the selected ontologies. The rest of the steps involve choosing user-defined functions (e.g., operators such as finding distance between two physical locations), and setting run-time parameters for the information request. This tool has been built over the IScape Builder in our previous version of InfoQuilt [STP01]. The IScape query optimization and execution details are discussed in our previous work [PS01].

6. Conclusions and Future Work
In this paper we have described a platform for creating, exchanging and sharing of knowledge so that we can obtain more useful information. We provide one initial step towards addressing the above challenge in two parts: bringing semantic capabilities of defining complex relationships, inter-ontological relationship, and knowledge discovery in the Semantic Web framework, and investigating how P2P can help in knowledge sharing. The key issues we have discussed include knowledge discovery based on semantic searching for relevant ontologies, and a peer-to-peer infrastructure in realizing ontology sharing and dissemination of knowledge on the Semantic Web. A prototype implementation has been completed for the capabilities discussed in this paper, and several examples can be demonstrated now.

As a future work to realize a more complete P2P semantic web vision, we plan to address following research issues:

![Figure 2: Example of Semantic Search and Exploration of Inter-Ontological Relationships](image-url)
• Discovering “Who Knows What?” – forming communities of interest, and designating leader (or knowledgeable) peers,
• Speaking the Same Terminology – enabling semantic interoperability among peers through conflict resolution, and semantic consensus generation,
• Asking a Question – hypothesis generation, and propagation including selective propagation through communities and “knowledgeable” peers.

References

http://www.daml.org/2001/03/daml+oil.daml#Thing.
