A Peer-to-Peer System to Share Ontology in the Semantic Web

Antonio M. Rinaldi
Dipartimento di Informatica e Sistemistica, Università di Napoli Federico II
80125 Via Claudio, 21 - Napoli, Italy
amrinald@unina.it

ABSTRACT
The problem of defining efficient techniques for knowledge representation (KR) is becoming more and more a challenging topic in both academic and industrial community. The Semantic Web needs of formal knowledge representations for implementing useful services and put in common the different views between man and machines. In this framework, we assume that new approaches for knowledge definition and representation may be useful, in particular the ones based on the concept of ontology. In this paper we propose novel techniques and tools for ontology definition and management. In particular, we propose a suitable model for representing knowledge and a system based on peer-to-peer (P2P) paradigm to share general and domain knowledge. The network knowledge is exposed using a Web Service and it can be used by intelligent Web agents.

Categories and Subject Descriptors
H.3.1 [Information Storage and Retrieval]: Online Information Services—Data sharing, Web-based services; H.4.m [Information Systems]: Miscellaneous

General Terms
Management, Design

Keywords
Ontology, Semantic Network, WordNet

1. INTRODUCTION
The large amount of information on the internet and the needs of implementing a formal definition of knowledge representation require a more efficient and effectiveness techniques for mining the Web. In the last years some new approaches strictly related to the above matter have been proposed and one of them is the Semantic Web, defined as: “an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation” [5]. In this context, the concept of relevance about a given topic has a basic importance. In our study we are interested to the definition of relevance given by Schutz [27] according to him, it is defined as the aboutness of information to a theme namely to the specific aspect or object of our concentration, giving as base an horizon that is the stock of knowledge at hand. Therefore we must define models and techniques to represent and manage this knowledge. New techniques have been developed to solve those problems. Some of them are based on ontologies to delete or at least smooth conceptual or terminological mess and to have a common view of the same information. The ontological aspects of information are intrinsically independent from information representation, so the information itself may be isolated, recovered, organized and integrated with respect to its content. A formal definition of ontology is proposed in [19] where conceptualization refers to an abstract model of a specific reality in which the component concepts are identified; explicit means that the type of the used concepts and the constraints on them are well defined; formal refers to the ontology propriety of being “machine-readable”; shared refers to the fact that an ontology captures the consensual knowledge, accepted by a group of persons. We also consider other definitions of ontology found in [24]. This definition indicates the way to proceed in order to construct an ontology: i) identification of the basic terms and their relations; ii) agreement on the rules to arrange them; iii) definition of terms and relations between concepts. From this perspective, an ontology includes not only the terms that are explicitly defined in it, but also those that can be derived using defined rules and properties. Thus an ontology can be seen as a set of “terms” and “relations” among them, denoting the concepts that are used in a specific domain. Moreover a central vision of the Semantic Web is that a user can cooperate with software applications using shared data and semantics. In this context intelligent applications interact with human agents to provide services. The ontological aspects of data need a hard work for formalizing its intrinsic knowledge. The ontology creation is a complex and hard consuming task therefore it is very important that these ontologies can be accessed and reused by more and more users and applications. According to [7] a distributed approach to Knowledge Management is useful for organizations and people which want to share “private knowledge”. Moreover, from a technological point of view a P2P approach is particularly well suited, because it make possible for different participants (organizations, individuals, or departments) to maintain their own knowledge structure while exchanging information. Knowledge sharing is a complex task; in fact it entails both knowledge creation and knowledge reuse. These two activities are not orthogonal, as new knowledge builds on the re-use of existing knowledge. With this aim, knowledge sharing has three main tasks: (i) location of relevant explicit knowledge; (ii) selection of relevant/significant knowledge; and (iii) application of the knowledge in a particular context. In this context the
notion of a community of practice [29] is of significant relevance. This approach is based on the idea that the knowledge can not be divided from the communities which create it. In our vision we consider as a central point the formalization of a knowledge representation in a community. This community can build ontologies using formal models and techniques and share them in a P2P environment. In this paper we propose a novel logical model to define ontologies and we use a standard language to represent them. Using this model we develop domain ontologies and we share them in a P2P network; with Web Service technology the network can show its knowledge to external agents for reusing. The paper is organized as follow: in the section 2 some related works are presented and discussed; in section 3 the system architecture and its main functionalities are described; in section 4 we define and describe our model using OWL; a case study is presented in section 5; conclusion and future works are in section 6.

2. THEORETICAL BACKGROUND

In this section we introduce some works related to our domain of interest in order to put in evidence the differences with our proposed method. In particular, we describe several systems and frameworks for knowledge sharing and approaches to formalize the knowledge using ontologies.

2.1 Systems

In the last years several approaches and systems have been presented to manage and share knowledge and in particular ontologies. In [1] is presented InfoQuilt, a system for sharing ontologies in a peer-to-peer environment. Using this system a user can find relevant sets of ontologies, reuse them, create new ones and advertise the resulting ontologies. The system allows to search concepts and services exploring inter-ontological relationships. The implemented system is based on agents that allow users to request information, semantically correlate data from different sources and of heterogeneous type or representation; they can have an interactive interface for knowledge discovery. Becker at al. [4] propose a P2P extension for ontology editing based on Ontorama [14]. The system uses a sharing protocol using RDF. This approach provides a novel editing environment compared to the classic client/server ontology management approaches. In [3] a method to integrate different knowledge sources as thesaurus, gazetteer and a chronology in an ontology using Topic Maps is proposed. This ontology is shown to Government Agencies by Web Services to support information harmonization about environmental data. KAON [13] is an open-source software infrastructure to manage ontologies for semantic-driven applications. It integrates traditional technologies as relational databases with new knowledge representation tools. There are several additional components which increase the KAON functionalities. COE (Cooperative Ontology Editor) [18] is a P2P application designed to allow ontology developers to share their knowledge through many activities: ontology sharing, ontology reuse and other traditional peer-to-peer mechanisms. It is implemented over COPPER, a framework for creating flexible collaborative P2P applications that provides nonspecific collaboration tools as plug-ins. SWAP (Semantic Web and Peer-to-Peer) [20] is a project that allows participants to keep private knowledge structures in their personal computer and share that knowledge in a P2P architecture. Users can extract ontologies from selected remote repositories, which are automatically integrated in their local repository. Any change in the source of the information is propagated to the local repositories. Oyster [26] is a java-based system, which assists users in managing, searching and sharing ontology metadata in a peer-to-peer network; it is a P2P application that exploits semantic web techniques in order to provide a solution for exchanging and re-using ontologies. The Oyster client on its own (e.g. disconnected from the P2P network) provides added value to its users as it will give researchers an overview and search facilities of his/her own ontology metadata. In order to provide this functionalities, Oyster implements a proposal for a metadata standard, so called Ontology Metadata Vocabulary (OMV) which is based on discussions and agreement in the EU IST thematic network of excellence Knowledge Web as the way to describe ontologies. In [23] the authors propose an integrated agent system for ontology sharing on WWW, which enables users to manage ontologies and Semantic Web Services. The proposed system has several modules to manage personal information, translate them into standard language as RDF and analyze RDF to obtain user’s interests and create Semantic Web Services which enable agent program to make inferences from grounding data on personalized ontology. In this context we notice also XAROP [11], a P2P platform to knowledge management in a decentralized IT infrastructure. Several surveys and books have been presented in the last years to evidence the importance of Peer-to-Peer and ontologies for enabling the Semantic Web; an useful reference is [28].

2.2 Models and languages

From a generic point of view we can define a generic model for knowledge representation as composed by a triple \(<S, P, C>\), where: \(S\) is a set of objects; \(P\) is the set of properties used to link the objects in \(S\); \(C\) is a set of constraints on \(P\). One of the most important progress in the KR applications derived from proposing [22], studying [30],[8],[9] and developing [10],[17],[6] of languages for knowledge representation. Even if those languages have several differences they share some common aspects based on the specification of objects (concepts) and the relationships among them. The main features of all KR languages are: Object-orientedness: all information about a specific concept is stored in the concept itself (on the contrary, e.g., of rule-based systems). Generalization/specialization: these properties are basic aspects of the human cognition process [22]; the KR languages have mechanism to cluster concepts into hierarchies where the high-level concepts represent more general attributes than the low-level ones which inherit the general concepts attributes but are more specific, presenting additional features on their own. Reasoning: the capability of inferring the existence of information not explicitly declared by the existence of a given statement. Classification: given an abstract description of a concept, there are some mechanisms to determine if a concept can have this description. This feature is a special form of reasoning. Object orientation and generalization/specialization help human users in understanding the represented knowledge; reasoning and classification guide an automatic system in building a knowledge representation as the system knows what it is going to represent. There are many points in common and several differences between our system and approach from the ones described before; in particular we propose a complete framework to manage, share and reuse ontologies. In addition our system has several features, such as editing functionalities and Web Service to share ontologies outside the P2P network. The several peers in our system have a general knowledge base as a support for knowledge organization together with a common model based on linguistic properties to build ontologies in order to have a common schema and a standard language for knowledge representation.

3. SYSTEM ARCHITECTURE

The proposed system has several software modules and, form a top-level view, they can be organized around some entities and...
argue from figure 1, a peer has two main tasks: (i) managing and editing local ontologies and (ii) putting in share local ontologies. The Rendez-Vous peer has a list of active peers and a description of their contents. It uses these information in the knowledge discovery step both between peers and in the sharing phase with Web Service. The Presence Advertiser advertises in the communication steps. In the following sub-sections we describe into details both the remaining modules drawn in figure 1 and the algorithm used to build dynamically the semantic network.

3.1 Ontology Extraction and Editing

In our approach we use an appropriate algorithm to extract from WordNet [21] (the system knowledge base) a domain semantic network; this net provides a general representation of our domain of interest. Many information systems use a knowledge base to represent data in order to satisfy information requests and in our vision it is a good choice having a common view of the same general and specific knowledge domain. Moreover in our framework WordNet can be a “starting point” for users because they can extract an initial general ontology from this knowledge base and expand it to have a specialized one; these tasks are explained in the following of this section. We implement our ontology by means of a semantic network. This structure is often used as a form of knowledge representation: it is a graph consisting of nodes which represent concepts and edges which represent semantic relations between the concepts. If the users domain of interest is not in the Ontology Repository, they can build an ontological domain using WordNet. This step is performed by the OntoExtractor which is in charge of extracting an ontology (i.e. a semantic network). We propose a dynamic construction of the semantic network using an ad hoc algorithm which takes into account the WordNet structure. WordNet organizes the several terms using their linguistic properties. Moreover, every domain keyword may have various meanings (senses) due the proprieties of polysemy, so a user can choose its proper sense of interest using the tool interface. Beyond the synonymy, we consider other linguistic proprieties applied to the typology of the considered terms in order to have a strongly connected network. Our network is built starting from a domain keyword that represents the context of interest for the user. We then consider all the component synsets and construct a hierarchy, only based on the hyponymy property: the last level of our hierarchy corresponds to the last level of WordNet one. After this first step we enrich our hierarchy considering all the other kinds of relationships in WordNet (see table 1). Based on these relations we can add other terms in the hierarchy obtaining a highly connected semantic network. Clearly, even if a knowledge base could be large and detailed, it will never give us a high level of specialization for every existing knowledge domains. Our approach tries to give a solution to this problem. In fact users can interact with our system in order to create a first ontological knowledge representation or they can expand it or create a new one using the OntoEditor module. Using the OntoEditor functionalities a user can modify the ontology structure as a whole adding new terms and concepts in the network, linking terms and concepts using arrows (lexical and semantic properties), deleting nodes and arcs. All the ontologies can be exported in OWL following a schema model described in section 4. The interaction with the semantic network using the editing tools is archived by means of Java 3D libraries. The algorithm to extract the semantic network from WordNet is described in pseudo-code in figure 2.

3.2 Advertisements and Sharing

The other main functionality of the peer system is the ability of sharing knowledge using ontology. This system feature is obtained sharing knowledge using ontology. This system feature is obtained by means of Java 3D libraries. The algorithm to extract the semantic network from WordNet is described in pseudo-code in figure 2.

Figure 1: The system architecture

Figure 2: Semantic network extraction algorithm
4. THE ONTOLOGY MODEL

The system has a common model for defining ontologies. In this way all peers have a common view of distributed knowledge and can share it in a simple way. The adoption of this model can help external agents (outside the P2P network) to use the ontologies for their purpose. In our approach the knowledge is represented by an ontology implemented by a semantic network. In this context we consider words, the defined properties are the linguistic ones and the constraints are the roles of validity applied on linguistic properties with respect to the considered term category. A concept is a set of words which represent an abstract idea.

Figure 3: Concept and Word

In the last years several languages have been proposed to represent ontologies; one of the most used is OWL. We use the DL (Description Logic) version of OWL because it has enough effectiveness to describe the ontology. The DL version allows the declaration of disjoint classes which are used, for example, to assert that a word belongs to a syntactic category. Moreover it allows the declaration of union classes used to specify domains and property ranges to relate concepts and words belonging to different lexical categories. Every node, both concept and word, is an OWL individual. The connecting edges in the ontology are represented as ObjectProperties.

Table 1: Properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Domain</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>hasWord</td>
<td>Concept</td>
<td>Word</td>
</tr>
<tr>
<td>hasConcept</td>
<td>Word</td>
<td>Concept</td>
</tr>
<tr>
<td>hypernym</td>
<td>NounsAndVerbsConcept</td>
<td>NounsAndVerbsConcept</td>
</tr>
<tr>
<td>holonym</td>
<td>NounConcept</td>
<td>NounConcept</td>
</tr>
<tr>
<td>entailment</td>
<td>VerbWord</td>
<td>VerbWord</td>
</tr>
<tr>
<td>similar</td>
<td>AdjectiveConcept</td>
<td>AdjectiveConcept</td>
</tr>
</tbody>
</table>

The use of domain and codomain reduces the property range application. For example the hyponymy property is defined on the set of nouns and verbs; if it is applied on the set of nouns it has as range the set of nouns, otherwise if it is applied to the set of verbs it has as range the set of verbs. In table 2 there are some of defined constraints and we specify on which classes they have been applied w.r.t. the considered properties; the table shows the matching range too.

Table 2: Model constraints

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Class</th>
<th>Property</th>
<th>Constraint range</th>
</tr>
</thead>
<tbody>
<tr>
<td>AllValuesFrom</td>
<td>NounConcept</td>
<td>hyponym</td>
<td>NounConcept</td>
</tr>
<tr>
<td>AllValuesFrom</td>
<td>AdjectiveConcept</td>
<td>attribute</td>
<td>NounConcept</td>
</tr>
<tr>
<td>AllValuesFrom</td>
<td>NounWord</td>
<td>synonym</td>
<td>NounWord</td>
</tr>
<tr>
<td>AllValuesFrom</td>
<td>AdjectiveWord</td>
<td>synonym</td>
<td>AdjectiveWord</td>
</tr>
<tr>
<td>AllValuesFrom</td>
<td>VerbWord</td>
<td>also_see</td>
<td>VerbWord</td>
</tr>
</tbody>
</table>

Sometimes the existence of a property between two or more individuals entails the existence of other properties. For example, being the concept dog a hyponym of animal, we can assert that animal is a hypernym of dog. We represent in OWL this characteristics by means of property features shown in table 3.

Table 3: Property features

<table>
<thead>
<tr>
<th>Property</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>hasWord</td>
<td>inverse of hasConcept</td>
</tr>
<tr>
<td>hasConcept</td>
<td>inverse of hasWord</td>
</tr>
<tr>
<td>hypernym</td>
<td>inverse of hypernym, transitivity</td>
</tr>
<tr>
<td>holonym</td>
<td>inverse of holonym, transitivity</td>
</tr>
<tr>
<td>entailment</td>
<td>cause</td>
</tr>
<tr>
<td>similar</td>
<td>symmetry and transitivity</td>
</tr>
</tbody>
</table>
5. THE CASE STUDY

We describe and test our methods and techniques with a complete use case in order to put in evidence the several features of the proposed model and implemented system. The system has been completely implemented using Java. We remind that the P2P network is obtained using JXTA libraries while the Web Service uses the AXIS framework. The P2P network is built starting from the first peer which creates a group and defines itself as Rendez-Vous Peer. The JXTA framework uses internal advertisements to represent network resources (i.e PeerGroup, pipe and services). When a peer wants to enter in a group an advertisement specifying some information about it must be sent to the Rendez-Vous Peer. When it wants share resources, other types of advertisement are sent and collected by Rendes-Vous Peer. We define a personalized advertisement managed by the Presence Advertisement module to have a more punctual description of our ontologies. These advertisements, called Ontology Advertisements, are stored in the Rendes-Vous Peer cache. They are used for ontology discovery in the P2P network and arranged in a list for the Web Service to expose the available ontologies out to the P2P network. Using the windows tabs shown in figure 5, a user can analyze the network and peer activities checking the log file in the Main Tab, share ontologies exploring its folders, create Ontology Advertisements, search and download ontologies using advertisements metadata.

![Figure 5: Peer Management Interface](image)

A user interacts with the system editing tools using an ad hoc interface is shown in figure 6. We choose as motivating example the building of ontology in territorial planning research field and in particular about rural landscape.

![Figure 6: Editor Interface](image)

This choice is motivated by the cultural and scientific innovations, which led in recent years to a completely new interpretation of landscape and it represents an hot topic in that community for its intrinsic complexity and by a collaboration of the author in a EU COST (European Co-operation in the field of Scientific and Technical Research) Action: COST21 “TOWNTOLOGY: Urban Ontologies for an improved communication in UCE projects”. The ontology is arranged following a methodology defined by a collaboration with a group of researchers at the Department of Planning and Territorial Science of the University of Napoli Federico II. In the ontology pre-consensus step has been defined a glossary using well-know knowledge sources as [12, 16, 15, 25], providing a more detailed description of the variables, since it was affected by several particular situations about the field of interest. A specific ontology about rural landscape has been created ex nvo using this glossary. The related elements and phenomena have been individualised by applying disaggregating and re-aggregating processes to the rural landscape components. The rural landscape image has been described by means of its natural (ecological) factors, built-up (settling) factors and visual and perceptive factors. All these macro-categories have been arranged in classes, which in turn have been divided in features and variables. Successively a domain ontology about landscape is extracted from WordNet following the steps described in the section 3.1 starting from the keyword region(sense3): a large indefinite location on the surface of the Earth. Even if we have many terms in this ontology it does not have a good detail level. In fact the landscape concept is in a low level in the WordNet hierarchy and few concepts are in the ontology representation; rural landscape concept not is in WordNet. The process to extract an ontology from WordNet starts from an interaction with the user inserting a specific term (i.e. region) by means of user interface and choosing the proper sense reading the description of the related concepts. The system gets the right sense and build the ontology following the steps described in the previous section. We choose to link the new proposed concept of landscape (an area, as perceived by people, whose character is the result of the action and interaction of natural and/or human factors [16]) directly to the region synset, using it as bridge for ontology merging. Therefore we expand this ontology giving more information about the rural landscape (see figure 6). Using our tool we have the OWL representation shown in figure 7. We notice that we give a knowledge expansion of WordNet knowledge base inserting new synsets about the territorial planning research field. The built ontology is shared in the network by peer owner using the Peer Management Interface described before.

![Figure 7: Owl representation of rural landscape ontology](image)
6. CONCLUSIONS AND FUTURE WORKS

The sharing and reuse of existing ontologies is a non-trivial task. A P2P approach allows the creation of knowledge communities in which information can be shared and reused in an effective way. Moreover the specialized knowledge in the local communities should be used by other sentient agents to perform their tasks in a more accurate way. On the other hand the needs of formal ontologies for knowledge representation involve several aspects of knowledge engineering and sharing systems. In this paper we have proposed a global approach to solve these problems; we implemented a system for sharing and creating ontologies in a P2P network, the network knowledge is exported out using a Web Service. Our system uses a simple and general model for knowledge representation taking into account a linguistic approach considered as the natural communication way between human agents. The ontologies are represented using OWL. We want to point out that, from a general point of view, we have an evolution of concepts during time; this is a cause of knowledge obsolescence, so there is the need for a continuous updating. In WordNet, for example, the concept of landscape is related only to a visual appearance dimension. Moreover in all fields of knowledge the research innovation allows the definition of new concepts. For example the concept of rural landscape does not exist in WordNet; this lack needs of a knowledge expansion.

Other issues need to be investigated, such as a solution to the ontology mismatch problem and the definition of an algorithm for automatic ontology merging to integrate peer domain ontologies in general knowledge bases used by intelligent agents on the Web.

7. REFERENCES


