

Ontology-based Web Knowledge Management

Yanmei Wang, Zhonghua Yang, Pe Hin Hinny Kong, and Robert Kheng Leng Gay
Information Communication Institute of Singapore (ICIS),
School of Electrical and Electronic Engineering,
Nanyang Technological University
SINGAPORE 639798

Abstract

The knowledge management is becoming more and more important in organizations, either over the intranet or Internet. In this paper we present an ontology-based web knowledge management (KM) framework based on web ontology language DAML+OIL. This framework supports content-oriented rather than traditionally document-oriented approach to knowledge management. Three fundamental building blocks, i.e., *annotations based on ontologies*, *knowledge-bases based on assertions in ontologies* and *web resources crawling*, and *rule-based reasoning/inference systems* for semantic knowledge manipulation. Our approach to knowledge management is the result of our semantic web research efforts. We adopt the web-standard based tools in our development of knowledge management. We believe that this approach of annotation-crawling-inference (A-C-I) to knowledge management is flexible and effective in supporting knowledge sharing on the web. An ongoing prototype is briefly described.

Keywords Ontology, DAML, DAML+OIL, Knowledge Management, Semantic Web

1. Introduction

The recent and growing interest in the Semantic Web has given rise to a flurry of activity in standardization body, W3C, to specify semantics using formal languages and inference mechanisms. As pointed by the inventor of WWW, Tim Berners-Lee, one of important application areas supported by the semantic web is knowledge management. Web-based knowledge management provides attractive alternative to the traditional knowledge management approaches. Organizations need knowledge management tools that enable users to better understand each other's changing contextual knowledge, and that foster effective and efficient collaboration while capturing, representing and interpreting the knowledge resources in their domain and business context [1].

Traditionally, much of Knowledge Management (KM) has focused on information technology, but has not lain out in clear terms what notions of "knowledge" need to be supported and to be accessed via standard web technologies. For example, IT-supported KM solutions are

built around some kind of organizational memory that integrates informal, semi-formal and formal knowledge in order to facilitate its access, sharing and reuse by members of the organization for solving their individual or collective tasks [2]. Knowledge management manipulates knowledge contained in documents. Somehow Knowledge can be acquired or "discovered" automatically from disparate, heterogeneous information sources, e.g., Web pages and networked document collections. However, this approach seems at best naive as it ignores the context and intended purpose of source information [18]. Without establishing this context and purpose, it seems unlikely that much useful "knowledge" can be discovered, as it is open to broad interpretation. Furthermore, the vast majority of existing document management system use keyword matching as a search method, combined with information retrieval without concerning the meanings for the intended search.

The semantic web, based on semantic knowledge representation enables communication between machines. The core idea of the Semantic Web is to make information accessible to human and software agents on a semantic basis [4]. Knowledge has to be modeled, structured and interlinked for supporting its flexible integration and its personalized presentation to the user. The challenge here is not data, but the rules and meanings about data (metadata) defined precisely enough so that machines can correctly interpret and quickly process that data as managed knowledge. Ontologies have shown to be the right answer to these structuring and modeling problems by providing a formal conceptualization of a particular domain that is shared by a group of people in an organization [3]. Ontology-based knowledge management use ontology as a backbone to represent metadata for providing and accessing knowledge sources. Ontology is the key to a semantic web.

In this paper, we present our initial efforts in Semantic Web. We explore the standard technologies from W3C, particularly the web ontology language DAML+OIL for knowledge management. An ontology-based framework is presented in which we adopt the approach based on annotation, crawling and inference. The idea of ontology-based knowledge management on the web is demonstrated using a prototype.

This paper is organized as follows. In Section 2, we provide a brief overview of emerging semantic web and web ontology languages. Section 3 describes our framework of

web knowledge management. In Section 4, the prototyping of concepts of proof is described. We discuss some related work in Section 5. We conclude this paper in Section 6.

2. Semantic Web and Ontology

As defined by Tim Berners-Lee *et al*, the Semantic Web is an extension of the current web in which information is given well defined meaning, better enabling computers and people to work in cooperation [19]. W3C is establishing standards for semantic web based on RDF. It is expected that a semantic web serves as a universally accessible platform that allows data to be shared and processed by machines as well as people. The RDF is a technique for representing knowledge. RDF networks consist of nodes with links between them. The nodes in a semantic network represent concepts. Links are paired to represent a relation and its inverse relation [5].

Ontologies, which capture the semantics of information and giving them a concise, uniform and declarative description, have gained significance due to the semantic web efforts. Ontologies embody human knowledge via symbols that are machine processible. Therefore, ontologies benefit machine reading the data when we use ontologies to index data and to express the metadata [6]. In the context of knowledge management, the ontology describes information based on its semantic context, rather than at the syntactic context, enables agents and computer software components to understand the semantic meaning of the keyword defined in ontology. For example, if we define the concept "Tiger" is an "Animal" and has property "has 4 legs". Next time when we use "Tiger", the semantic meaning is implied that "this animal Tiger has 4 legs". When we search the semantic web, software agents or software components capture Tiger's semantic meaning once "Tiger" concept is captured. After the query is submitted to the search engine, the correct results are returned (Figure 1).

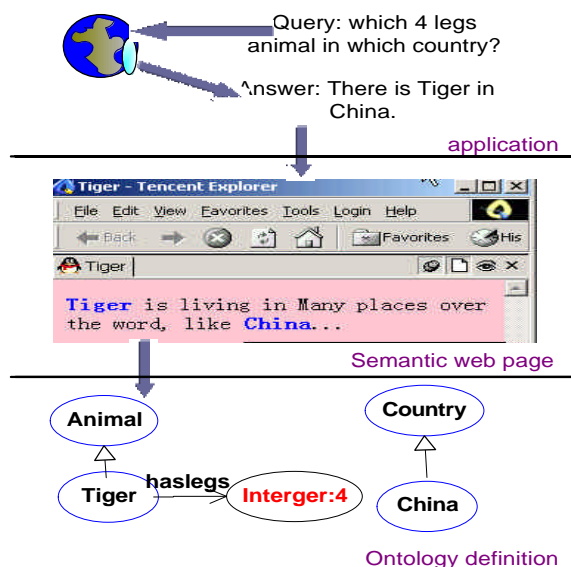


Figure 1: Ontology and Semantic Query

Web ontologies are specified in a web ontology language. A number of languages for defining ontologies on the Web, such as RDF(S), DAML+OIL, and OWL, are under development. These ontology languages incorporate the best knowledge representation practices available in the industry and are flexible and extensible enough to easily incorporate new representational features and incorporate and interoperate with different knowledge models. RDF schemas differ from XML schemas in that they do not define syntax but instead classes, properties, and their interrelation: they operate directly at the data model level, rather than the syntax level. Scaled up over the Web, RDF schemas are a key technology, as they allow machines to make inferences about the data collected from the web [7].

The DAML+OIL language has been edited by the joint committee for the US DAML (DARPA Agent Markup Language) and European OIL (Ontology Inference Layer) projects and emphasis is put on the work done to support the W3C XML Schema datatypes [11]. DAML allows one to be even more expressive than with RDF Schema, and brings us back on track with our Semantic Web discussion by providing some simple terms for creating inferences [7].

3. A Framework of Web Knowledge Management

The skill to effectively and efficiently manage knowledge is vital for sustainable corporate success. Ontology-based technologies might be the key to the of knowledge management applications [8]. Most importantly, ontology-based knowledge management enables knowledge reuse and knowledge sharing by share a common understanding.

In this section, we present a framework to implement ontology-based web knowledge management application. The framework has three levels for information and knowledge flow. They are annotation, crawling, and inference of knowledge. We call it A-C-I Approach (Figure 2).

3.1 Knowledge Flow

The first step to design a knowledge management system is to identify the knowledge flows of the system. Knowledge flows comprise a set of processes, events and activities through which data, information, knowledge and meta-knowledge are captured, transferred, and transformed from one state to another [10]. Our framework has four types of knowledge artifacts due to different level of conceptualization..

In order to capture and represent knowledge, we develop domain ontologies after consulting the domain experts. Annotators will embed in semantic concepts modeled by ontology into their web pages. In an essence, the web resources (typically web documents) are marked in a web ontology language. Basically, all assertions and instances as specified in ontologies and those instances marked in web

documents will go into knowledge base (KB). The knowledge base is an integral part of in reasoning (inference) system. The reasoning system applies rules to derive new knowledge.

To get the knowledge artifacts in from the unstructured knowledge, we use knowledge processes, they are annotation with ontology development, crawling, inference (Figure 2). Annotation and development process derive the unstructured knowledge to structured knowledge with semantic meanings. Crawling process gathers the dispersed knowledge in semantic web pages. Inference process makes the knowledge machine processible and intelligent. As a result, the above three processes are also based on ontology and are the fundamental building blocks of our web knowledge management framework.

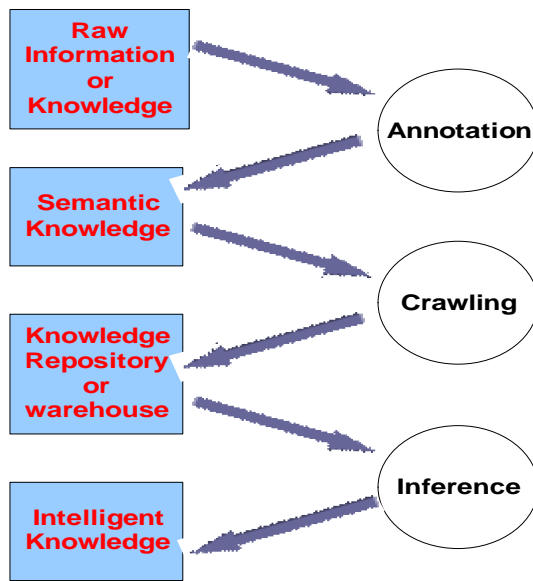


Figure 2: Knowledge Flow with Tools Used

3.2 Ontology-based Knowledge Management

Knowledge management involves knowledge base used to store knowledge items and relations among knowledge items. The knowledge base is the absolute collection of all expertise, experience and knowledge of those within any organization [9]. As indicated previously, our knowledge base is closely related to ontologies and reasoning systems.

Ontology-based knowledge management here is an application that offers structured access to large volumes of unstructured information in the Internet and intranet. Ontology serves as the foundation for ontology based knowledge management to allow sharing knowledge over the web.

From analysis of the knowledge artifacts and knowledge processes in the knowledge flow, we have developed a KM framework as shown in Figure 3. This framework supports the knowledge flow as identified above, and has three major functional units for annotation and ontology development, crawling, and inference. In addition,

DAMLJessKB is used to add some inference rules to knowledge base [15]. The specific use of these tools will be discussed in some detail when we discuss the prototyping. Our framework is flexible enough for supporting developing any ontology, any application, and instance data embedded in the annotated web pages in a specified domain to manage the knowledge on the web.

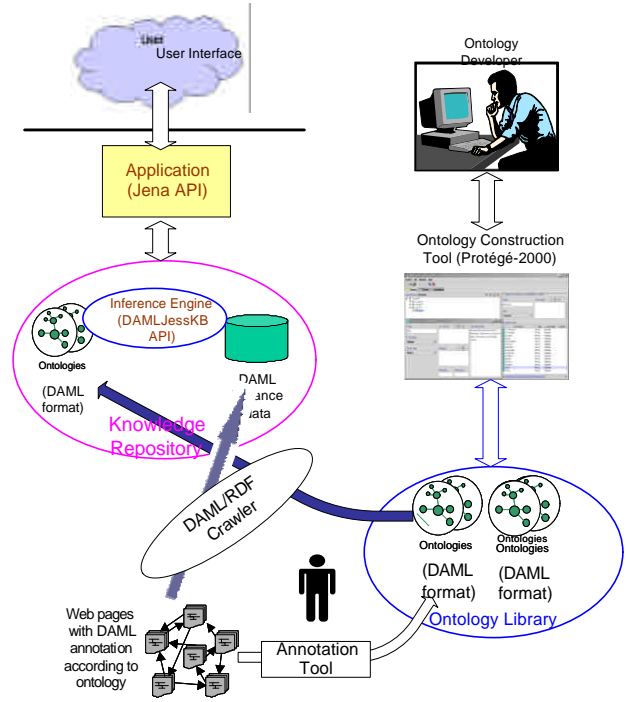


Figure 3: A-C-I Framework

3.3 Ontology Development and Annotation

We use Protégé-2000 as ontology and knowledge base editor. Protégé is the easiest to use and supports the construction of knowledge-acquisition interfaces based on ontological data [20]. We add a DAML+OIL plugin in order to support our ontology and instance creation.

The annotation tool supports the task of creating and maintaining ontology-based DAML+OIL markups i.e. creating of DAML-instances, attributes, relationships, and enriching their web pages with DAML-meta data [21].

3.4 Crawling

The building of knowledge base relies on RDF crawler [12]. RDF crawler downloads interconnected fragments of RDF from the Internet and builds a knowledge base from this data. RDF crawler can gather DAML statement as well as DAML+OIL based on RDF-Schema. To enable embedding in other tools, RDF Crawler provides a high-level programmable interface (Java API). RDF Crawler utility is just a wrapper around this API - either a console application, or a windows application or a servlet. The detailed specification of RDF crawler is referred to [12]. For our purpose, we are interested in its capability to process

DAML+OIL. It helps get ontologies from the ontology library and gather RDF statement and DAML+OIL assertions from the web resources.

3.5 Inference

The crawling results are sent to a knowledge base which in effect containing ontologies and instances. This knowledge base combining with the assertions (rules) forms the core of ontology-based knowledge management, the latest knowledge artifacts in knowledge flow (Figure 2). The inference engine, in conjunction with knowledge base, answers queries based on semantics. Furthermore, it derives new knowledge by a combination of facts and the ontology [13]. Our reasoning system is a rule-based system based on DAMLJessKB which allows adding rules to knowledge repository [15]. DAMLJessKB utilizes the Java Expert System Shell (Jess [14] is a rule engine and scripting environment written entirely in java language) developed at Sandia National Laboratories to carry out this reasoning.

3.6 Toolkit for KM Application Development

As seen from Figure 3, our framework includes toolkits for developing KM applications. While the framework does not mandates the use of specific toolkits or development environment, we found that HP Labs Semantic Web toolkit Jena is a convenient toolkit to manipulate RDF models for developing applications within the semantic web [22,16]. The Jena API is designed specifically for the Java programming language. The application developed using Jena API will process the instance data according to the metadata in ontology. In the developer's side, a domain expert is consulted to get the domain knowledge developing good ontologies.

4. Prototype

This model of annotation-crawling/inference guides us in developing various KM applications. In this on-going research, we are developing a prototype for demonstration purpose. We have defined the application as a simple skill management system for human resource department in an IT company. The application and the ontology developer may implement other kinds of application and other domain ontology according to specific needs and business domains. There are many researches on the ontology development.

We have developed the skill ontologies, which conceptualize and describe an IT company human resource skill structures. A snapshot of the ontologies is shown in Figure 4. A piece of codes following shows Skill ontology representation using DAML+OIL as follows where the name space is "xmlns:daml_oil=<http://www.daml.org/2001/03/daml+oil#>".

Note that there are related concepts concerning "Person" which models details of employees and applicants in the company. The layout of other ontologies is shown here due to length restriction.

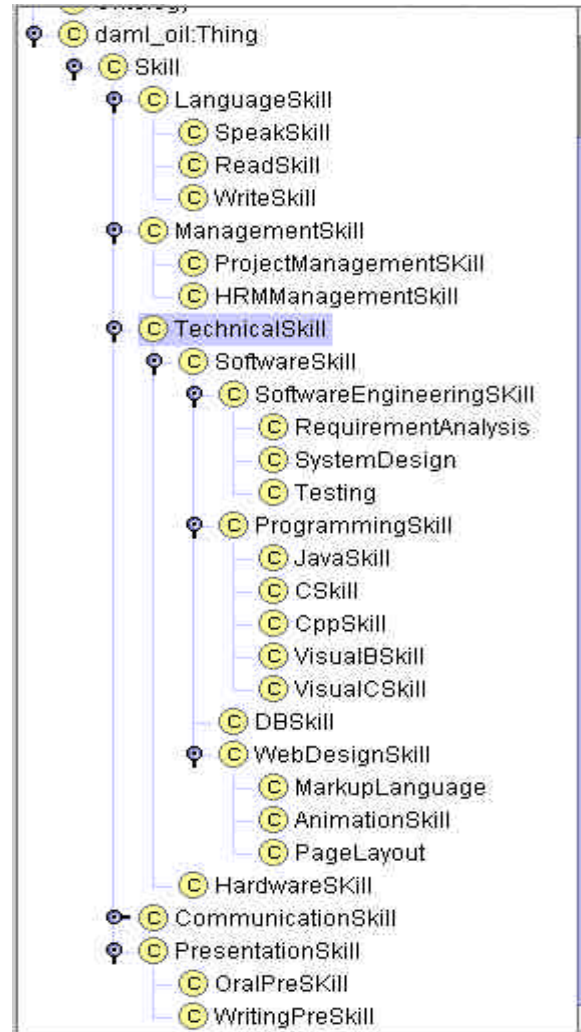


Figure 4: Skill Ontology

```
<daml_oil:Class rdf:ID="Skill">
</daml_oil:Class>
<daml_oil:Class rdf:ID="TechnicalSkill">
  <rdfs:subClassOf rdf:resource="#Skill"/>
</daml_oil:Class>
<daml_oil:Class rdf:ID="SoftwareSkill">
  <rdfs:subClassOf rdf:resource="#TechnicalSkill"/>
</daml_oil:Class>
<daml_oil:Class rdf:ID="SoftwareEngineeringSkill">
  <rdfs:subClassOf rdf:resource="#SoftwareSkill"/>
</daml_oil:Class>
<daml_oil:Class rdf:ID="Testing">
  <rdfs:subClassOf
rdf:resource="#SoftwareEngineeringSkill"/>
</daml_oil:Class>
<daml_oil:DatatypeProperty rdf:ID="skillLevel">
  <daml_oil:domain rdf:resource="#Skill"/>
  <daml_oil:range
rdf:resource="http://www.w3.org/2000/10/XMLSchema#integer"/>
</daml_oil:DatatypeProperty>
```

5. Related Work

AQSearch is a Software Engineering project (Northeastern University College of Engineering) to develop a software

application that supports the process of storing and retrieving project information. This information is stored as DAML-annotated documents. Appropriate ontologies capturing the domain knowledge (in this case the domain is software engineering and project management) are part of the deliverables. Various parties then may view this information by posting queries through a query engine. The engine retrieves information and presents the answers to the users [17].

The AQSearch is not enough in practical application over an organization, especially an international organization being splitted over the world. AQSearch put information (instance data) in the DAML file, same file extension as ontology file. In practical, the information maybe in the web pages, in the database, or somewhere else. Our A-C-I approach overcomes the problems of AQSearch and add some more functionalities in. the framework presented search from the web through crawler, not only the daml document specified. That is to say, this model is more complete than AQSearch.

6. Conclusions

In this paper we have shown our approach of ontology-based web knowledge management. From our point of view, our proposal is rather unique because it is web-based, easy to implement, and layered architecture due to knowledge artifacts.

For the future, we see a number of new important topics appearing on the horizon. For instance, we consider study of ontology conversion among different ontology representation language. It helps ontology more wide ontology reuse and human resource saving.

7. Acknowledgements

The research presented in this paper would not have been possible without sponsor of EEE School and NTU-Sun ASP Centre.

References

- [1] York Sure, Steffen Staab, Rudi Studer. Methodology for Development and Employment of Ontology based Knowledge Management Applications. Institute AIFB, University of Karlsruhe. Association for Computing Machinery, Volume 31, Number 4. December 2002.
- [2] R. Dieng, O. Corby, A. Giboin, and M. Ribiere. Methods and tools for corporate knowledge management. Int. Journal of Human-Computer Studies, 51(3):567-598, 1999.
- [3] D. O'Leary. Using AI in knowledge management: Knowledge bases and ontologies. IEEE Intelligent Systems, 13(3): 34-39, May/June 1998.
- [4] Nenad Stojanovic, Alexander Maedche, Steffen Staab, Rudi Studer, York Sure. SEAL-A Framework for Developing Semantic PortALs. Lecture Notes in Computer Science, Institute AIFB, University of Karlsruhe. 2001.
- [5] Edward Pattison-Gordon, Knowledge Representation for Medical Concepts.
- [6] Jeremy Rogers. Developing and applying ontologies: experiences from the medical domain. Medical Informatics Group, University of Manchester. 2002.
- [7] Sean B. Palmer. The Semantic Web: An Introduction. SGML Center Presentation. UK. October 2001.
- [8] Christian Ohlms, Mckinsey & Company. The Business Potential Ontology-based Knowledge Management. KnowTech 2002. 15 October 2002.
- [9] Aegiss Corporation. Introduction to Knowledge Management and the Knowledge Base. 1995.
- [10] Brian (Bo) Newman. A Framework for Characterizing Knowledge Management Methods, Practices, and Technologies. January 1999, in support of The Introduction to Knowledge Management, George Washington University Course EMGT 298.T1, Spring 1999.
- [11] Eric Van Der Vlist. DAML and OIL embrace W3C XML Schema. 28 Mar 2001 UTC.
- [12] Kalvis Apsitis. Specification of a RDF Crawler. Institute AIFB, University of Karlsruhe.
- [13] Steffen Staab, Jurgen Angele, Stefan Decker, Michael Erdmann, Andreas Hotho, Alexander Maedche, Hans-Peter Schnurr, Rudi Studer, York Sure. AI for the Web — Ontology-based Community Web Portals. Institute AIFB, University of Karlsruhe. In AAAI 2000.
- [14] Ernest Friedman-Hill. Jess: The rule engine for the Java platform. <http://herzberg.ca.sandia>. 1995.
- [15] Joe Kopena William C. Regli. DAMLJessKB: A Tool for Reasoning with the Semantic Web. Geometric and Intelligent Computing Laboratory, Department of Computer Science, College of Engineering, Drexel University. October 28, 2002
- [16] Brian McBride. Jena: Implementing the RDF Model and Syntax Specification. Hewlett Packard Laboratories. Bristol, UK. 2000.
- [17] J. H. Tu, R. Feng, Z. Y. Li, W. Tong, Z. Q. Liu. MIM3152 & MIM3153 PROJECT SPECIFICATION. Northeastern University College of Engineering. January 07, 2002 – June 06, 2002.
- [18] Clifford Behrens and Vipul Kashyap. The “Emergent” Semantic Web: A Consensus Approach for Deriving Semantic Knowledge on the Web. In the Processings of the Firs Semantic Web Working Symposium, pp 55-74, July 30-1 August 2001.
- [19] Tim Berners-Lee, James Hendler, Ora Lassila. The Semantic Web, Scientific American, May 2001.
- [20] Protégé Users Guide, http://www.smi.stanford.edu/projects/protege/doc/users_guide/index.html
- [21] What is OntoMat-Annotizer? <http://annotation.semanticweb.org/tools/ontomat>
- [22] HP Labs Semantic Web Research, <http://www.hpl.hp.com/semweb/index.html>.