Design and Implementation of Semantic Community Web Portal *

Ching-Long Yeh and Chang-Gang Chen
Department of Computer Science and Engineering
Tatung University
40 Chungshan North Road, Section 3
Taipei, 104, Taiwan
chingyeh@cse.ttu.edu.tw g8906003@mail.ttu.edu.tw

Abstract

In this paper, we employ the RDF technology to build a semantic web portal. The portal is divided into content layer and program layer. The contents of the portal are represented in RDF according to the schema defined in ontologies. The program layer consists of application and indexing programs. The application programs includes conceptual search, directory navigation, forum service and personalization service. The indexing programs consist of ontology editor, RDF parser, and content extraction program. The separation of content and program makes it easy to reuse of program for different domains. At present we take library as the domain of this research. Our next step is to reuse the program layer by replacing content of another domain.

1 Introduction

HTML is the main technology used to create the content of web sites. It is a good document standard for presentation and web navigation. It, however, suffers greatly from being difficult to access and maintain the content. XML [1] separates structure and presentation in document and it improves to some extent the problems of HTML. However, the XML structure, either in DTD [1] or Schema [2], can only represent the compositional relationship between the higher and lower level elements, which is not sufficient as a representation of semantic content in document. RDF [3], a data model in directed label graph, enhances the shortcomings of XML by allowing user expressing the property values of subjects. An RDF document thus can be viewed as collection of subject-predicate-object statements. An RDF schema is a class of RDF having a specific vocabulary for nodes and edge labels. RDF Schema (RDFS)[4], a special schema, provides facilities for user to define class, subclasses, property, etc., which is used as the basis for ontology development. DAML [5], adding practical considerations, is an extension of RDFS and becomes the most important ontology language. Based on the common knowledge of ontology, the semantic content of RDF files created by different parties can then be shared, processed, integrated, and reused by machines. The Semantic Web is a vision towards this [13]. In this paper, we attempt to develop a web portal using the ontology technology.

The essential idea is that all the information in the portal, either newly created or imported from sources in other format, is represented in RDF according to ontologies obtaining from web site. Then we develop a number of portal services, including conceptual search and directory service based on domain conceptual hierarchy, agent-based personal services like news filtering and event notification, and forum service. We choose library domain as the target of the design and implementation of the web portal.

This research is carried out in two parts. First is

*This research was supported by the Taiwan National Science Council under Contract No. 91-2422-H-036-322.

the preparation of semantic content for the portal. The sources of content are mainly from (1) the annotation of documents in various formats like HTML and PDF, (2) importing data from SQL database, (3) user's profiles, and (4) discussion data in forum. We therefore need the following ontologies as the schema of the above semantic content: publications (subclassified as book, magazine, and journal), author, user's profile, time, news, events, and forum. We did not construct the ontologies from scratch; instead we collected from ontology library, such as DAML ontology library, and made necessary modifications for our purpose. For the task of annotating document, we employed OilEd as the annotation tool. The task of importing data from SQL database, mainly for library online catalogue, is concerned with getting data in database tables and stored in RDF according to the schema of publication ontology. The sources of items 3 and 4 listed above are obtained when subscribing in the portal and posting in a forum of special interest. We provide graphical user interface for user's convenience to enter the above data, which is then stored in RDF. The second part is concerned with the design and implementation of application services for accessing the web portal. In this paper, we provide user with conceptual search, directory service, new filtering, event notification and forum service.

In Section 2 we give a brief review of the background of the research. In Sections 3 and 4, we describe the design and implementation of the web portal. Finally, we make conclusions and describe future work.

2 Relevant Background

RDF

The Resource Description Framework (RDF) is a general framework for the description of Internet resource such as a Web site and its content. An RDF description (such descriptions are often referred to as metadata, or "data about data") can include the authors of the resource, date of creation or updating, the organization of the pages on a site (the site map), information that describes content in terms of audience or content rating, keywords for search engine data collection, subject categories, and so forth. The RDF will make it possible for everyone to share Web site and other descriptions more easily and for software developers to build products that can use the metadata to provide better search engines and directories, to act as intelligent agents, and to give Web users more control of what they're viewing.

Table 1: Web Pages was created by Richard and a number "123456"

Object	**Attribute**	**Value**
http://www.w3.org/	creator	#pages
#pages	name	Richard
#pages	number	123456

RDF's basic data model is simple: besides resources, it contains properties and statements. A property is a specific aspect, characteristic, attribute, or relation that describes a resource. A statement consists of a specific resource with a named property plus that property's value for that resource. This value can be another resource or a literal value: free text, basically. Altogether, an RDF description is a list of triples: an object (a resource), an attribute (a property), and a value (a resource or free text). For example, Table 1 shows the three triples necessary to state that a specific web page was created with a name "Richard" and a number "123456." You can easily depict an RDF model as a directed labeled graph. To do this, you draw an oval for every resource and an arrow for every property, and you represent literal values as boxes with values. Figure Three-triples shows such a graph for the triples in Table 1. These example notations reveal that RDF is ignorant about syntax; it only provides a model for representing metadata.

Ontology

Ontologies are explicit formal specification of the terms in the domain and relations among them [15]. They are used to define common concepts, properties, relationships, and constraints for people to share understanding in a domain. Ontologies range from very

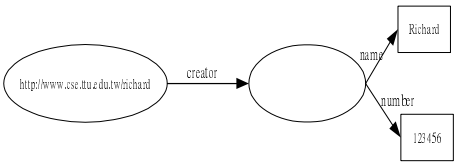

Figure 1: The diagram for the triples in Table 1

general terms that form the foundation for knowledge representation in all domains, to terms that are restricted to specific knowledge domains [16]. In recent years the development of ontologies has been moving from the field of Artificial-Intelligence laboratories to the knowledge management [22] and the Web Services integration [18]. In meta-ontology level, many standards are introduced, such as KIF[19], RDFS[4], DAML+OIL[20]. Methodologies of ontologies engineering are also available in many researches [21].

Web Portals

A web portal is a web site that collects information for group of users having common interests [14]. A web portal provides interfaces for user to locate interested information in the portal according to the directory structure, personal interest, topic, *etc.* Users of common interests can build their own community to submit and share information.

To help locating information, index of information is provided either by member of group under some topic or tagged by the content provider. The indices become the metadata to identify the topic of the content. However, the subject-based indexing is sufficient to satisfy the request by members of the community portal. Thus it is necessary to index the content of web portal with deeper semantic information in order to make the services effective. To make web portal meaningful, web portal can define ontologies for the community. The ontologies provide expressive schema for describing content of portal. Inference based on the schema can be used to improve the access of portal content.

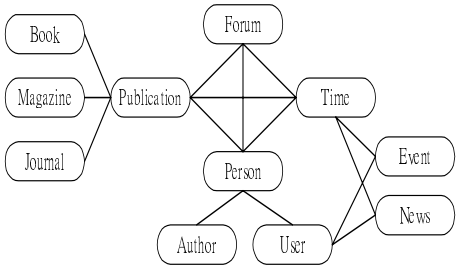

Figure 2: The Relationships among ontologies

3 System Design

We refer to the knowledge processes describe in [22] to formulate the following steps to build the portal.

1. **Importing ontologies:** Build or choose ontologies as the schema of metadata.
2. **Annotation:** Annotate the content of portal according to the ontologies.
3. **Indexing:** Create indices for the metadata obtained in the preceding step.
4. **Use:** The indexed metadata is ready to be used by the application programs.
5. **Maintenance:** The process loops back to the import step if new ontologies are created.

In the following we describe each step in turn and then conclude with a system diagram.

Importing ontologies

This step is concerned with preparing ontologies to be used as the schema for the description of portal content. In this paper, we decide to select appropriate ontologies for our purposes instead of creating new ones. The source of ontologies is mainly from DAML Ontology Library [23]. Since our current domain is library, we built ontologies for describing publications, person, forum, time, news and events. The publication ontology is further sub-categorized as book, magazine and journal; the person ontology has two sub-ontologies, author and user. The relationships among the ontologies are summarized as shown in Figure 2.

The ontologies are imported in an annotation tool, here Ontoedit [24], as the schema to be used in Annotation Step when creating metadata.

Annotation

The Annotation Step is to annotate the portal contents with metadata. We use Ontoedit as the annotation tool to create metadata. As mentioned in the previous step, we have imported the ontologies into Ontoedit, which results in a class hierarchy. The portal contents need to be annotated are (1) documents in various formats like HTML and PDF, (2)data from SQL database, (3) user's profiles, and (4) discussion data in forum. The first type of documents include news and events. For each document of this type, we create an instance of either news or event class. The annotator then fills in the value of each property of the new instance as shown in the user interface of the annotation tool. After finishing the task, we then export and save the instance as an RDF file. Book, journal and magazine information belongs to the second type of content. Since the information is stored in tables, we therefore extract data from the tables and store them in an RDF file of the corresponding schema using SAX APIs in Java [25]. The third type of metadata describes the personal information, including the interested subjects, schedule, and contact information. A profile is created for a person when subscribing the portal. User enters the personal information through a form-based interface. After submitting, the information is stored in an RDF file according to the schema of user profile. The last type of content is created in a way similar to the third type. The portal site provides a form-based interface for user to post their opinions. The fields in the form are arranged according to the schema of the forum ontology. The data collected from the interface is therefore stored in an RDF file using the forum schema.

Indexing

The task of Indexing step is to parse the RDF files collected in the Annotation Step and then store in a relational database [1]. We use an RDF parser, SiRPAC [26], to convert RDF files into the corresponding triple statements. We create a table to store the resulting triples. For each ontology we create a table closely related to the schema of the ontology. Each triple is store as a record in the triple table, and the data in the triple is stored in the record of the corresponding table.

[1]We choose relational database because of its efficiency and reliability.

Use

This step is concerned with providing services for user to access the content of portal. In this paper, we provide user with conceptual search, directory service, new filtering, event notification and forum service. Both the conceptual search and directory services are built based on the conceptual hierarchies in the ontologies. The conceptual search interface provides user with specifying query as the target class along with the attribute-value pairs as the constraints of the class. The search engine then searches the RDF instances that satisfy the conditions and return the result to user. The directory service interface lists the top classes found in ontologies, so that user can click on the interest class to expand the detailed content and list the associated instances. User can then choose instances of interest. The portal provides the following agent-based personal services: new filtering, event notification, and book reservation. News filtering and event notification services notify user by email or Short-Message-System with selects news of interest and new events. Finally, the forum service provides places for users to exchange opinions for specific topics. It is a GUI that allows user to enter data for each field without knowing the class hierarchy behind the interface. The resulting data collected from the interface is stored in RDF. We then use the conceptual search and directory service described previously to help user access the content. Furthermore, the forum service provides personal services such as subject-based article filtering by combining user's profile and the forum RDF.

After the prototype is built, we collect comment from user to improve the functions of the portal. When enhancing or modifying the service applications, the ontolgies may need to be modified or new ones are created. We then loop back to the first step.

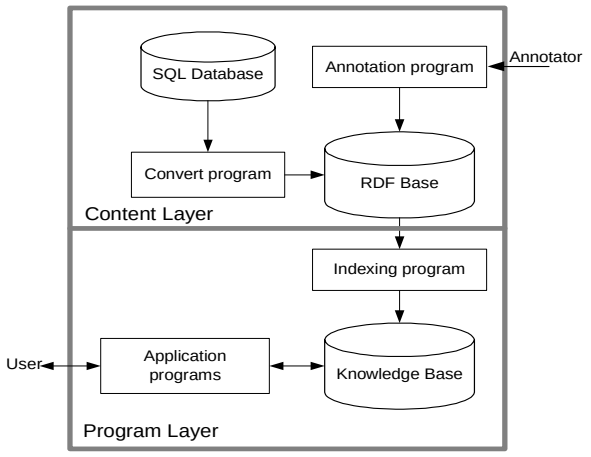

Figure 3: Conceptual architecture of the web portal

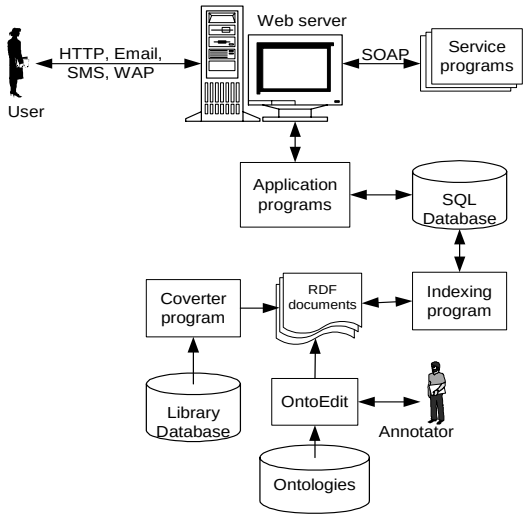

Figure 4: System Architecture of the library web portal

Conceptual architecture

The steps described previously is summarized as a two-layer conceptual architecture as shown in Figureărchitecture. The content layer corresponds to the importing and annotations steps described previously, while the program layer corresponds to the indexing and use steps.

4 Implementation

We follow the procedure described previously to build a library web portal by using the content of Tatung University Library. The system architecture is shown in Figure s̃ystem-architecture). The application programs, including conceptual search, directory navigation, forum service, news filtering and event notification, and login program are implemented using HTML and JSP technology. The web server consists of two programs Microsoft IIS and Apach Tomcat to serve the interaction between user and system. At present, the interactions are made through HTTP and email. We are trying to extend the way of interaction using mobile devices, for example, WAP and SMS. In addition to the application programs we also have implemented the programs of the content layer as shown in the lower part of the system architecture. We use OntoEdit as the annotation tool, imported with three ontologies, event, news and time, to create metadata for the event and news documents created in the library site. The converter and indexing programs are implemented using Java with SAX and SiRPAC. The result of indexing is stored in a SQL database that is used as the knowledge base of the application programs. There are existing programs for library services, for example, booking service, expiration notification, that are integrated into the web site using SOAP.

Portal site demo

When a new user enters the site, she is asked to input personal information, including contact information, schedule and subjects of interest. After successfully login the site, user is presented with a list of services provided by the application programs. User can select the services of interest. The service of event notification, news filtering and forum service will present personalized information, as shown in Figure 5, for user according to the information in the profile.

Figure 5: The personalized page

5 Conclusion and Future Work

In this paper, we describe the design of a semantic web portal. We have implemented a prototype of library web portal. It achieves most of the primitive functions described previously. This research is a pilot study of our ontology-based application of digital archive project. Its mission is building a web portal using the ontology technology, so that we can see how it works, the strong and weak points of ontology-based services. Comparing with conventional web portals, we can make better use of the semantic content in the RDF document to build new higher-level services instead of creating new HTML document and building database to account for new applications.

In this research we employed relational database technology to build the knowledge base for the use of application programs that were implemented using Java technology. Due to the lack of inference capability supported by both technologies, it is difficult to add inference rules when developing the service programs. In another paper [**?**], we have developed a RDF triple store using relational database technology and Prolog. Prolog is a rule-based language, which supports rich inference capability. The inference capability can be used to complement the weak point of this research. Our next step is extending our experience obtained from this experiment to the field of digital museums.

References

[1] W3C, Extensible Markup Language (XML), Available at http://www.w3.org/TR/REC-xml

[2] W3C, XML Schema Part 0: Primer, W3C Recommendation, 2 May 2001, Available at http://www.w3.org/TR/xmlschema-0/

[3] W3C, Resource Description Framework (RDF), Available at http://www.w3.org/RDF/

[4] W3C, RDF Schema (RDFS), Available at http://www.w3.org/TR/rdf-schema/

[5] DAML Ontology Library, Available at http://www.daml.org/ontologies/

[6] Alexander Maedche and Steffen Staab, Ontology learning for the Semantic Web, *IEEE Intelligent Systems*, Volume: 16, Issue: 2, pp. 72-79.

[7] S. Staab, M. Erdmann, and A. Maedche, Ontology in RDF(S), *Electronic Articles in Computer and Information Science* ISSN 1401-9841, Vol. 6(9), 2001 Available at URL: http://www.ep.liu.se/ea/cis/2001/009/

[8] S. Staab and A. Maedche, Knowledge Portal – Ontologies at Work, *AI Magazine*, Volume 22, number 2, pp. 63-75.

[9] R.Scott Cost, et al., ITtalks: A case stidy in the Semantic Web and DAML+OIL, *Proceedings of Semantic Web Working Symposium (SWWS) 2001*, pp. 477-494.

[10] J. Heflin, R. Volz, J. Dale, W3C Working Draft 07 March 2002, Requirements for a Web Ontology Language,

[11] A. Maedche, et al., SEAL –A Framework for Developing Semantic Web PortALs, *BNCOD 2001 - 18th British National Conference on Databases.*, pp. 1-22.

[12] N. F. Noy and D. L. McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology*, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

[13] http://www.w3.org/2001/sw/

[14] W3C, Requirements for a Web Ontology Language, W3C Working Draft 07 March 2002, http://www.w3.org/TR/webont-req/.

[15] T. T. Gruber. 1993. A Translation Approach to Portable Ontology Specification. *Knowledge Acquisition*. 5: 199-220.

[16] B. Chandrasekaran, J.R. Josephson, V.R. Benjamins. 1999. What Are Ontologies and Why Do We Need Them? *IEEE Intelligent Systems*, Volume 14, No.1, pp.20-26.

[17] S. Staab, R. Studer, H. Schnurr and Y. Sure (2001). Knowledge Processes and Ontologies. *IEEE Intelligent Systems*, Volume 16, No. 1, pp. 26-34.

[18] S. McIlraith, T.C. Son and H. Zeng. 2001. Semantic Web Services. IEEE Intelligent Systems, *Special Issue on the Semantic Web*, Volume 16, No. 2, pp. 46-53.

[19] Knowledge Interchange Format,draft proposed American National Standard (dpANS), http://logic.stanford.edu/kif/dpans.html.

[20] DAML+OIL- the revised language specification, http://www.daml.org/2001/03/daml+oil

[21] M. Uschold, and M. Gruninger. 1996. ONTOLOGIES: Principles, Methods and Applications, *Knowledge Engineering Review*. 11(2).

[22] Steffen Staab, Rudi Studer and Hans-Peter Schnurr. Knowledge Processes and Ontologies. *IEEE Intelligent Systems*, Jan/Feb 2001.

[23] DAML Ontology Library, Available at http://www.daml.org/ontologies/

[24] Ontoedit, Available at http://www.ontoprise.de/com/start_downlo.htm

[25] Java XML Pack, Available at http://java.sun.com/xml/downloads/javaxmlpack.html

[26] SiRPAC RDF Parser, Available at http://www-db.stanford.edu/ melnik/rdf/api.html

[27] Ching-Long Yeh and Ruei-Feng Lin, Design and implementation of an RDF triple store, *Proceedings of DATF-2002*.