Proceeding of the International Conference on Advanced Science, Engineering and Information Technology 2011

Hotel Equatorial Bangi-Putrajaya, Malaysia, 14 - 15 January 2011

ISBN 978-983-42366-4-9

The Semantic Portal for Supporting Research Community: a Review

Rahmat Hidayat, Yazrina Yahya, Shahrul Azman Mohd Noah

Knowledge Technology Research Group, Faculty of Information Science & Technology
Universiti Kebangsaan Malaysia, Bangi 43600 UKM Selangor Malaysia
E-mail: {rahmat, yaz, samn}@ftsm.ukm.my

***Abstract*— Current state of the art of typical search engines like Google, Yahoo and others are delivering references in terms of web URL or links to the related website. As such the results did not deliver the right answers required to the users needs. In addition to that as soon as the users require a collection of the information obtained, these search engines failed to do so resulting in the human intervention in-combining the information from several sources. Due ot the advancement and the vast number of sites and information on the web, demands in providing higher precision results are required to aid users in obtaining the most relevant result to the search process. One of the promising areas of the Semantic Web is enhancing the query capabilities for information. Small vertical vocabularies and ontologies have emerged, and the community of people using these and generating data is growing daily. However queries or search mechanisms that utilizea the vasrt amount of vocabularies, ontologies and data in digital libraries is still very much lacking. Therefore searching over heteregoneous records, data in digital library community or the Web has become a well known problem to the mass public. As such a solution is needed for a federated search across multiple resources available. However it remains unclear on how Semantic Web or its technology is used in constructing a digital library system or aid in enhancing the quality of the search results performed. This leads to the current work proposed, as work will be conducted to provide possible components that will construct the semantic web portal. The work performed is essential to facilitate semantic searches for research community in large-scale distributed digital library system. The subject research community is chosen particularly to aid in ensuring hat result obtained are accordingly to the users relevant needs. The expected outcomes of the research are an architecture that utilizes the semantic technology that will promote semantic web portal in the digital library and a semantic search mechanism that will provide better results and a combination of useful results relevant to the users.**

***Keywords*— research community, semantic portal, semantic search engine, ontology.**

I. INTRODUCTION

The portal website provides an environment for people from various backgrounds and experiences with the same interests, hopes and dreams to share. A website community (WC) can be seen as a web communication and social interaction in a group of people with common interests [1]. Web communities are commonly known as eCommunity, virtual communities, online communities, social networks or in the literature [2]. In one of his earliest works WCs, Dream of WC as a social phenomenon that does not have a business dimension [3]. Recent advances in information and communication technology has, however, the trip to WCs as a business enabler in the digital market provided. For example, Successful companies use virtual community, interaction between customers and suppliers to create products companies. As a community enables the exchange of experiences, problems and solutions discussed, established, and a sense of community among the members [4][5]. A virtual community of this type of economic activity such as e-commerce can be used.

A community may be referred to a group of entities that share similar properties or connect to each other via certain relations. The important goals of community minly can be seen as the process of identifying these connection and locating entities in different communities as well to have various applications. The application for finding potential collaborators for researchers by discovering communities in an author-conference social network, or recommending books (or other products) for users based on the borrowing

records of other members of their communities in a library system [6].

II. Development the Semantic Portal

Current Web technology experiences serious limitations to make information accessible for users in an efficient mechanism. The general problem to find information on the Web is summarized in : searches are imprecise, often yielding matches to many thousands of hits. Moreover, users face the task of reading the documents retrieved in order to extract the information desired. These limitations naturally appear in existing Web portals based on this technology, making information searching, accessing, extracting, interpreting and processing a difficult and it becomes a time-consuming task.

In this context, the Semantic Web enables automated the access of information [7] and its utilization based on machine-processable semantics of data. Ontologies are the backbone technology for the Semantic Web and - more generally - for the management of formalized knowledge in the context of distributed systems. They offer machine-processable semantics of data and information sources that can be communicated between different agents (software and people). In other words, information can be constructed understandable for the computer, thus it will give more opportunity for people to search, extract, interpret and process information.

Therefore Semantic Web technologies can considerably improve the information sharing process by overcoming the problems of current web portals. In this sense, portals based on Semantic Web technologies represent the next generation of web portals.

The essential aspect that is focused on this study can be seen as the state of the evolution of web portals and survey existing portals that make use of Semantic Web technologies [9]. The scope of portals investigated is restricted to Semantic Web portals (SW portal for short), which are defined as follows:

- It is a web portal. A web portal is a web site that collects information for a group of users that have common interests
- It is a web portal for a community to share and exchange information
- It is a web portal developed based on semantic web technologies.

Even with the methodological and tool support we have described so far, developing a Web portal for a community of non-trivial size remains a complex task. Strictly ad-hoc rapid prototyping approaches easily doom the construction to fail or they easily lead up to unsatisfactory results. Hence, we have thought about a more principled approach towards the development process that serves as means for documenting development, as well as for communicating principal structures to co-developers and editors of the Web portal. The different phases in the development process as seen in the Fig.1 can be seen as the first different feature. For the main part this model is a sequential one. Nevertheless, at each stage there is an evaluation as to whether and as to how easily further development may proceed with the design decisions that have been accomplished before. The results feed back into the results of earlier stages of development. In fact, experiences gained by running the operational system often find their way back into the system.

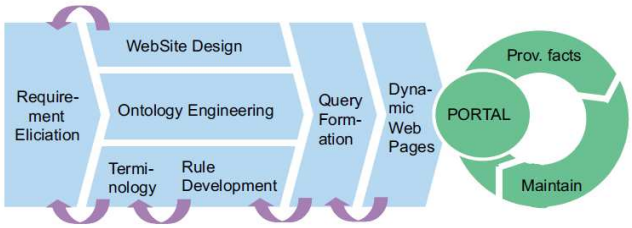

Fig. 1 The Development Process of the Community Web Portal

The main stages of the development process and their key characteristics are given in the following.

- The process is initiated with the elicitation of user requirements in the requirements elicitation phase. In this phase, requirements about important and interesting topics in the domain are collected, the information goals of potential users of the portal are elicited, and preferences or expectations concerning the structure and layout of presented information are documented. Results of this very first phase constitute the input for the design of the Web site and for preliminary HTML pages and affect the formal domain model embodied in the ontology.
- The requirements determine, e.g., which views and queries are useful for users of the portal, which navigation paths they expect, how different Web pages are linked, or which functionality is provided in different areas of the portal. Requirements like these are realized in the Web site design. This design phase may be performed independently to a very large extent from the underlying formal structuring, i.e. the ontology. Since a mock-up version of the Web site is developed early in the development phase, one may check early whether the system to be developed really meets the users' needs.
- In parallel to the development of the structure and layout of the Web site an ontology engineering process is initiated. The first phase elicits relevant domain terms that need to be refined and amended in the ontology engineering phase. First, the static ontology parts, i.e. the concept hierarchy, the attributes, and relations between concepts, are formally defined. Thereafter, rules and constraints are developed. Rule development may necessitate a major revision of the concept hierarchy. For instance, new subconcepts may have to be introduced, attributes may have to turn into relations or into other concepts, or relations may have to become concepts. This (intra-ontology) engineering cycle must be performed until the resulting ontology remains sufficiently stable.
- In the query formulation step the views and queries described in one of the earlier phases are formalized. At first, their functionality is tested independently from the Web site design. To express the information needs formally, the developer has to access the ontology, whereby additional rules or relations that

define new views or ease the definition of queries may become necessary. In order to test ontology and queries, a set of test facts has to be prepared. During this process of testing, inconsistencies in the ontology may be detected, which may lead to a feed back-loop back into the ontology engineering phase.

- Finally, Web pages are populated, i.e. the queries and views developed during Website design, and formalized and tested during query formalization are integrated into the operational portal.

Some semantic portals which have been developed by utilizing the SWRC ontology can be seen as follows:

- **AIFB** The portal of the Institute AIFB uses the SWRC to annotate staff, publications, projects and their corresponding interrelationships. The usage of the ontology will be explained below. http://www.aifb.uni-karlsruhe.de
- **OntoWare** Ontoware is a software development community platform for Semantic Web related software projects. The SWRC ontology is extended for annotating developers and software projects. http://www.ontoware.org
- **OntoWeb** The community site for the European Union founded project OntoWeb uses the SWRC to facilitate ontology-based information exchange for knowledge management and electronic commerce [10]. http://www.ontoweb.org
- **SEKT** The SEKT project is a European research projects bringing together the communities of knowledge discovery, human language technology and Semantic Web. Its corresponding portal uses the SWRC much like the OntoWeb portal. http://www.sekt-project.com/
- **SemiPort** The SemIPort (Semantic Methods and Tools for Information Portals) project develops innovative methods and tools for creating and maintaining semantic information portals (annotated by the SWRC ontology) for scientic communities. http://km.aifb.uni-karlsruhe.de/semiport

III. Ontology of Research Community

A community can be defined as a group of entities that share similar properties or connect to each other via certain relations. Identifying these connections and locating entities in diferent communities is an important goal of community mining and can also have various applications. The application for finding potential collaborators for researchers by discovering communities in an author-conference social network, or recommending books (or other products) for users based on the borrowing records of other members of their communities in a library system [6].

Representing knowledge about researchers, research communities, their publications and activities as well as about their mutual interrelations is a prime use case for distributed, locally maintained, interlinked and highly structured information in the spirit of the Semantic Web.

KA2 Project has started constructing ontology model to present research community. The SWRC ontology initially grew out of the activities in the KA2 project [1]. The SWRC ontology - initially phrased Semantic Web Research Community Ontology - which we will describe in this paper, generically models key entities in a typical research community and reflects one of the earliest attempts to put this usage of Semantic Web Technologies in academia into practice [8].

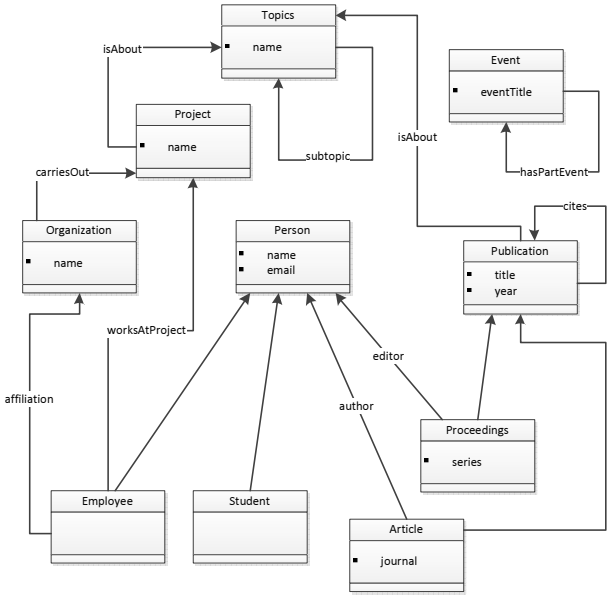

Fig.2 Main Concepts of the SWRC Ontology

SWRC comprises at total of six top level concepts, which can be identified as the Person, Publication, Event, Organization, Topic and Project concepts. Figure 1 presents a small portion of the SWRC ontology with its main top-level concepts and relations. The Person concept models any kind of human person and a large number of properties restrict their domain or range to individuals of this concept like student or author, respectively. The Person concept is specialized by a large number of - not necessarily disjoint - subconcepts, e.g. Employee, Student and the like. The Event concept is meant to model different types of events and is thus specialized by a wide range of concepts including events like Lecture or Conference. The Publication concept subsumes all different types of research publications modelled in close correspondence with the well known BibTEX publication types like Article or InProceedings. The Organization and Project concepts model more abstract concepts like the subconcepts Department or SoftwareProject, respectively. Both concepts can participate in a large number of relations like for example to the Person concept via the employs or member relations. The Topic captures arbitrary topics of interest which are arranged on the instance level via the subTopic relation. [8]

IV. Examples of the Ontology

As described in the previous explations, the SWRC ontology currently comprises seven class ontology (Document, Event, Organization, Person, Product, Project, and Topic). In the following, we show global overviews of

two sub-ontologies: the Person-ontology and the Document-ontology.

The Person-ontology model describes such concepts as person, the types of persons working in academic environments, along with their characterictics. This ontology consists of 19 classess.

Class hierarchy (19 class defined) :

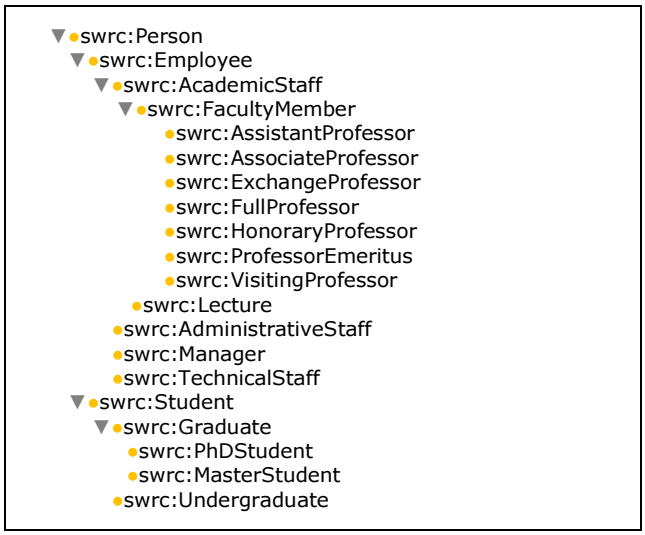

Fig.3 The Person-Ontology

The Document-ontology exposses such concepts as document which consists of 23 classes which can be seen in the figures 4.

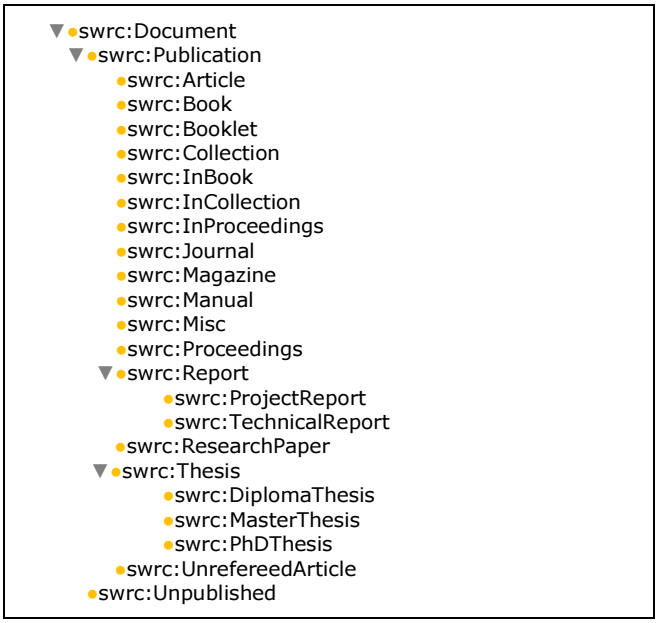

Fig.4 The Document-Ontology

V. Future Web Portal

The future development can be seen as the way how portal website can provide integrated discovery of life science research activities taking place accros university physical, administrative, and disciplinary divisions, which have been developed by Cornell University Library [11]. The system comprises a custom RDBMS-based knowledge base store and Java Web aplication driving the public portal as well as a Web-based ontology editing interface. The portal web offer browsing and searching by instance type, hyperlinked traversal of object, relationships and some simple inferences to aggregate data across multiple object properties.

VI. Conclusions

The SWRC ontology - The Semantic Web Research Community Ontology, which models key entities in research communities. The early research projects in Semantic Web Community can be seen at the original utilization of the SWRC ontology which has been consecutively refined up to the current OWL version of the ontology which we have presented in this paper and is thus a good example for ontology reuse. It is therefore a good choice for a ready and established out-of-the-box ontology in the research domain, thus e.g. saving time and effort for building a comparable schema.

Application of semantic web is very useful in the future to simplify the search of resource especially the resources that have close relationship with the research.

References

[1] Adiele C. Towards promoting interactivity in a B2B web community. Information Systems Frontiers. 2009. Available at: http://www.springerlink.com/index/10.1007/s10796-009-9187-7.

[2] Adiele, C. (2008). Modeling interactivity in a b2b web-based community. In IEEE International conference on digital information management (ICDIM 2008). Piscataway: IEEE.

[3] Rheingold, H. (2000). The virtual community: Homesteading on the electronic frontaier (revised edition). Cambridge: MIT.

[4] Bakos, Y. (1998). The emerging role of electronic marketplaces on the internet. Communications of the ACM, 41(8), 35–42.

[5] Hagel, J., & Armstrong, A. (1997). Net gain: Expanding markets through virtual communities. Boston: Harvard Business School Press.

[6] O.R. Za, J. Chen, and R. Goebel, “DBconnect : Mining Research Community on DBLP Data,” Network, 2007, pp. 74-81.

[7] H. Lausen, M. Stollberg, and R.L. Hernández, “Semantic Web Portals – State of the Art Survey,” World Wide Web Internet And Web Information Systems, 2000, pp. 1-47.

[8] Y. Sure, S. Bloehdorn, P. Haase, J. Hartmann, and D. Oberle, “The SWRC Ontology – Semantic Web for Research Communities,” Main, pp. 1-12.

[9] S. Staab, J. Angele, S. Decker, M. Erdmann, A. Hotho, and A. Maedche, “Semantic community Web portals,” Computer Networks, vol. 33, 2000, pp. 473-491.

[10] V.R. Benjamins and D. Fensel, “Community is Knowledge ! in (KA) 2 Feasibility of the approach and deliberative restric-,” Social Science, pp. 1-18.

[11] Lowe.B, Caruso.B and Corson.J.R, “Vivo Development Roadmap: Enhancing an Ontology-Based University Research Portal with OWL and Rules” Cornell Unversity, Mann Library, Ithaca NY, USA.