ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology

Vol. 6, Special Issue , August 2019
International Conference on Recent Advances in Science, Engineering, Technology and Management at Sree Vahini Institute of Science and Technology-Tiruvuru, Krishna Dist, A.P

Effective Information Retrieval in Current Information Systems

G Rasagna, B.Haritha, V.Divya, Y. Anusha

U.G. Student, Department of Computer Science&Engineering, Sree Vahini Institute of Science & Technology , Tiruvuru, A.P, India

ABSTRACT: In this paper we describe the functional requirements for research information systems and problems which arise in the development of such a system. Here is shown which problems could be solved by using knowledge markup technologies. In this article use one DAML and OIL ontology for Research Information System is offered. The already developed ontologies for research analyzed and compared. The architecture based on knowledge markup for collecting research data and providing access to it is described. It is shown how RDF Query Facilities can be used for information retrieval about research data.

KEY WORDS: Current Research Information System, Ontology, Information Retrieval, DAML, RDF, Knowledge Markup.

I. INTRODUCTION

Information about research results, projects, publications, organizations, researchers and so on published on the web play a more and more pervasive role in modern research. The increasing dependence of modern research on already achieved research results requires to have ability to retrieve research information in a more efficient way. Information overload by the exponential rise of amount of information makes it difficult for researchers to find relevant information. To solve these problems a number of Current Research Information Systems (CRIS) is being developed. But in most cases such systems do not solve their task of providing complete and actual information with a minimum of information noise. This is one reason that researchers are not prone to publish results about their research via information systems. Publishing usually is limited to researcher's or project's web pages.

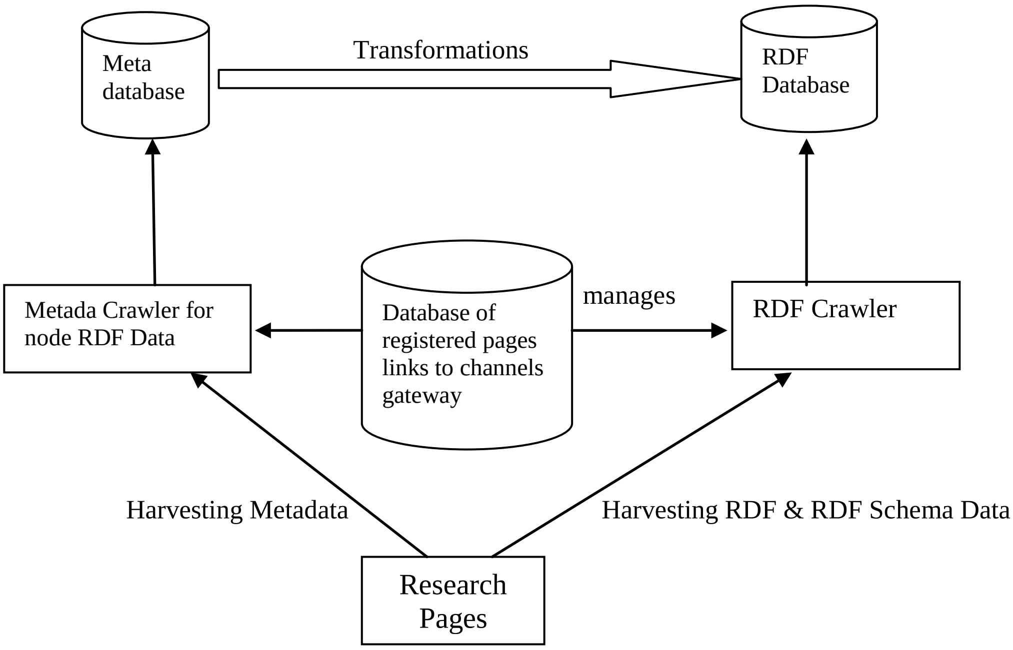
To provide actual and complete information for interested persons, information from research web pages also should be included into information retrieval operations. Usually researchers' or policy-makers' demands for research information is not limited to information from one single system. Research information in any science or technology area is scattered among a number of heterogeneous information systems. There is a strong need to gather information or to point researchers to systems where information can be found. It is very important to know if the gathered research information is actual and complete.

We are developing the AURIS-MM information system (Austrian Research Information System – Multi Media enhanced) to provide research information to interested consumers in a more attractive way. The system is being developed coming from the existing AURIS (Austrian Research Information System) and Fo Dok-Online (Research Documentation of Vienna University of Technology).
Our experience and newest web technologies showed us that centralized database systems are very efficient but not the best solution to provide access to research data due to a widespread distribution of the research data over the web.

ONTOLOGY DEVELOPMENT FOR SCIENCE

Some efforts already were done to provide to researchers, industry, policy-makers efficient information access to research data from some sectors of science and access to research limited to organization (university research information systems), or limited to geographical boundaries (national networks, ERGO[ERGO] – European Research Gateways Online) .

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Special Issue , August 2019
International Conference on Recent Advances in Science, Engineering, Technology and Management at Sree Vahini Institute of Science and Technology-Tiruvuru, Krishna Dist, A.P



Metadata collecting into RDF database

II. SIGNIFICANCE OF THE SYSTEM

Some efforts already were done to provide to researchers, industry, policy-makers efficient information access to research data from some sectors of science and access to research limited to organization (university research information systems), or limited to geographical boundaries (national networks, ERGO[ERGO] – European Research Gateways Online) .

III. LITERATURE SURVEY

Due to the fact that already huge amount of data is provided on internet web pages of projects, researchers, universities, it is hard to get researchers provide their data once more into a centralized system.But they can not limit search to trusted data, understand context of the page and provide search based on meaning of the data.One of the possible ways to collect data about research is the page annotation. Knowledge can be annotated on the page in such a way that automatic tools can collect and understand it. Ontologies make possible that software agents can understand knowledge which is marked. The benefits of ontologies and Semantic Web use for scientific publishing were described at [Lee-2001] Some effort is already done to develop markups for scientific data.

SHOE is a small extension to HTML which allows to annotate some knowledge about web page content. SHOE is a very simple language for declaring ontology, defining classification, relationship, inference rules, categories, etc. SHOE was developed in the Department of Computer Science, University of Maryland. SHOE specification, tools, SHOE ontology in plain text and DAML, examples are accessible at the SHOE home page. Several ontologies for university and research data were developed for SHOE.OIL (Ontology Inference Layer) [OIL, Fens-2000] - "is a proposal for a web-based representation and inference layer for ontologies, which combines the widely used modeling primitives from frame-based languages with the formal semantics and reasoning services provided by description logics. It is compatible with RDF Schema (RDFS), and includes a precise semantics for describing term meanings (and thus

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 6, Special Issue , August 2019

International Conference on Recent Advances in Science, Engineering, Technology and Management at Sree Vahini Institute of Science and Technology-Tiruvuru, Krishna Dist, A.P

also for describing implied information)." OIL was sponsored by the European Community via the IST projects Ibrow and On-To-Knowledge.

In the OIL for research data there were developed SWRC (Semantic Web Research Community ontology) (http://ontobroker.semanticweb.org/ontologies/swrc-onto- 2000-09-10.oil) and KA2 (Ontology of Knowledge Acquisition community). DAML (DARPA Agent Markup Language)[DAML] - ontology markup language, was developed as an extension to RDF and RDFS. DAML allows to specify ontologies and markup pages for automatic knowledge extraction. The last version of DAML is named DAML + OIL. DAML specifications, examples, tools, ontologies are published at DAML home page.

QUERYING COLLECTED METADATA, GETTING KNOWLEDGE FROM ANNOTATIONS

Once the annotated metadata were collected, how to use them? There are several tools which can be used to search annotated pages.SHOE Search Engine – Semantic Search http://www.cs.umd.edu/projects/plus/SHOE/search/) search registered annotated pages. User of search engine can choose ontology, then choose type of resource he searches, create very simple filter conditions and search SHOE metadata database. Our approach assumes that data would be described in RDF or can be translated into RDF by transformation procedure. Also to provide search services for researcher query facilities should be able to search data by its meaning (type of resource or property), values of attributes (properties) and relation between resources.
There are several query engines for RDF[Karv-2000], Squish, Ontobroker, Redland RDF Application Framework, MetaLog, RDF Data Query Language. In our project to query RDF database Sesame RDF Query Repository and Querying Facility is used. Sesame supports RQL (RDF Query Language) [Vass] which is being developed by ICS-FORTH Institute. Sesame supports storing both RDF and RDF Schema information. Querying Facilities of Sesame supports Schema information about subclasses and subproperties, searching by attributes values, resource relations.

WEB PAGE ANNOTATION

So the ontology can serve for understanding meaning of data. But to make data understandable by software agents, they should be provided in a format, which agent can parse A number of annotation tools are described in [Staab-2001]. For page annotation we use two tools: OntoMat and AURIS-MM metadata generating facilities. OntoMat [OntoMat] is a user-friendly interactive webpage annotation tool. It includes web browser and ontology browser. Ontology browser supports DAML + OIL ontology exploration. Web browser supports web browsing, highlighting parts of the web pages and creating annotations based on highlighted part of the pages. To annotate the web page researcher needs to open web page in the browser, then open ontology from provided by project URL. Then the researcher can crate annotation highlighting regions of the page and describing them in ontology browser according to the ontology terms, relation and attributes. OntoMat automatically creates RDF annotation and new web page with included RDF annotation. The annotated web pages can be published on the web instead of annotated.

IV. METHODOLOGY

ONTOLOGY

So, the main goal of our ontology development was to develop an ontology which will help users of research information to retrieve relevant information.
The Primary use cases of information retrieval for CRIS are [Jeff-98, CERIF-2000, Lind-2000, Aks-2000]

- Retrieving information about research results by researchers or students for results reuse. The estimation of research results.
- Seeking collaborators which can take part in research projects as partners, sell their expertise, results and intellectual rights
- Finding facilities and equipment which can be used for research
- Assess and access to Research and Development capabilities by policymakers

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 6, Special Issue , August 2019

International Conference on Recent Advances in Science, Engineering, Technology and Management at Sree Vahini Institute of Science and Technology-Tiruvuru, Krishna Dist, A.P

- Finding ongoing research and technology activities and results of projects by users in commerce and industry

Finding the sponsors for a new research project

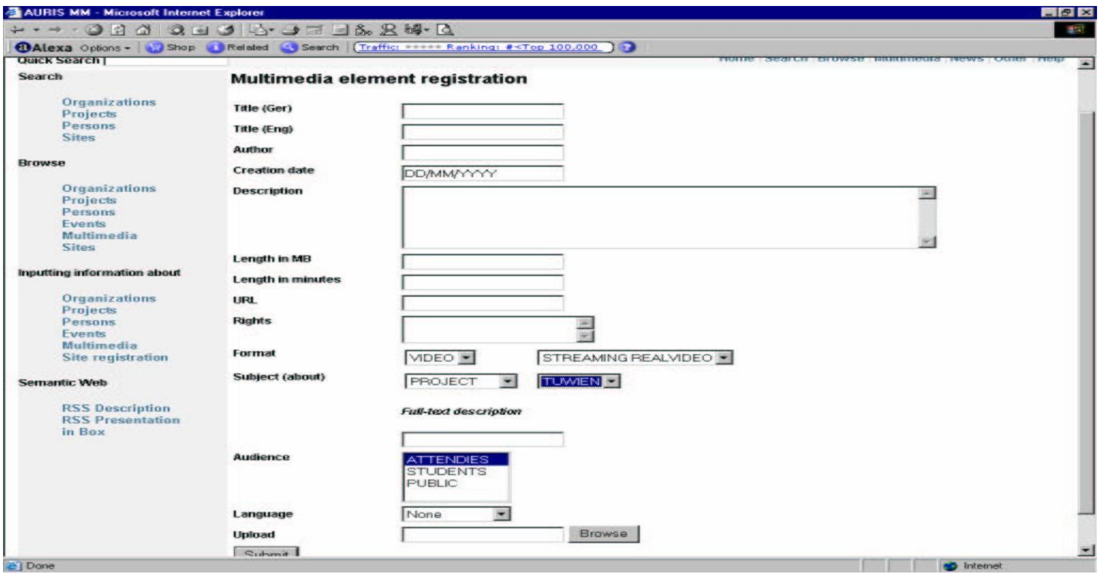
INFORMATIONAL RETRIEVAL ARCHITECTURE

The research data for retrieval should be collected, analyzed. To make possible analysis and understanding of meaning of data by software, they should be published in format understandable by software agent or annotated. Then annotations should be collected, analyzed, if it is considered necessary, they should also be transformed into one model/format. During search operation queries and data should be processed by search engines and response should be send to information consumers

So the process of information retrieval consists of

1. knowledge markup (by researcher)
2. harvesting marked-up knowledge by crawlers or software agents
3. transforming harvested data into formats appropriate for metadata repository/search engines
4. loaded into repository
5. retrieved by search engines according to users request

V. EXPERIMENTAL RESULTS



The registration of multimedia element.

Research results which can be reused might be described in publications (articles, thesis, technical reports, etc.). Research results might be described precisely (Research result or Product). They can be presented by advanced presentation techniques - Multimedia element, which maybe video, images, drawing, diagrams, MS PowerPoint presentations.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Special Issue , August 2019
International Conference on Recent Advances in Science, Engineering, Technology and Management at Sree Vahini Institute of Science and Technology-Tiruvuru, Krishna Dist, A.P

CERIF 2000	Math-Net ontology	SWRC Semantic Web Research Community	University Ontology
Person			
Yes. Not classified in CERIF	Yes.	Advanced hierarchy suitable for research and education	Advanced Hierarchy suitable for research and education
Project			
Not classified in CERIF	Yes	Yes. Classified.	No
Organization			
Yes. Classified	Yes	Close to CERIF Classification	Only educational

http://derpi.tuwien.ac.at/~andrei/cerif.rdfs#Person All persons in database (and any subtype of a person, -researchers and student)
http://derpi.tuwien.ac.at/~andrei/cerif.rdfs#Researcher All persons who are researchers (or any subtype of researchers)
^***http://derpi.tuwien.ac.at/~andrei/cerif.rdfs#Researcher*** All persons, who are researchers and not any subtype of researcher
select X,Y ***from #Project {X}. #project_persons{Y}, {Z} #expertise_skill {E}*** ***where X = Z and N = "Semantic Web"*** All projects in Semantic Web with description of persons participation in them If the organization or person, or Research Information System asserts new type of project – software project and in RDF Schema provides that it is a subtype of AURIS-MM, then it will also searched.
select X,Y ***from ^#Project {X}. #project_persons{Y}, {Z} #expertise_skill {E}*** ***where X = Z and N = "Semantic Web"*** Only projects in Semantic Web asserted as exactly CERIF

projects and participants of those projects

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 6, Special Issue , August 2019

International Conference on Recent Advances in Science, Engineering, Technology and Management at Sree Vahini Institute of Science and Technology-Tiruvuru, Krishna Dist, A.P

VI. CONCLUSIONS

Use of Semantic Web technologies might be very fruitful for development of Research Information Systems.

The annotation of knowledge makes it easier to researchers and research organization to assert information about their research for dissemination. No need to register it in a number of information systems. Software agents can collect information and understand its meaning

Not only research data but also new domain knowledge can be also asserted and shared for use.

Query engines for Semantic Web due to that inference abilities and schema exploration can make development of Research Information System more easy then conventional technologies like Relational Database management systems because exploration of domain knowledge is very crucial for CRIS systems.

REFERENCES

1.ERGO European Research Gateways Online http://www.cordis.lu/ergo
2.BL-2001 Berners-Lee T., Hendler J., Lassila O., The Semantic Web, Scientific American, May 2001 Hend-2001 Hendler J., Agent and the Semantic Web, IEEE Intelligent Systems Journal, March/April 2001
Erd-2001 M. Erdmann, A. Maedche, H-P. Schmurr, and S. Staab, From Manual to Semi-automatic Semantic Annotation, LIflk þIflg El9GtKOflIG AKtIGl9S Ifl CONþUt9K ðfld IflfOKNðtIOfl SGI9flG9, VOl 6 (2001)
3.SWA Semantic Web Activity http://www.w3.org/2001/sw Lee-2001 Berners-Lee T., Hendler J., Scientific publishing on the 'semantic web', 12 April, The Nature, http://www.nature.com/nature/debates/e-access/Articles/bernerslee.htm
4.Hefl-99 Jeff Heflin, James Hendler, and Sean Luke, SHOE: A Knowledge Representation Language for Internet Applications, Technical Report CS-TR-4078 (UMIACS TR-99-71). 1999. http://www.cs.umd.edu/projects/plus/SHOE/pubs/#tr99
5.SHOE SHOE home page. http://www.cs.umd.edu/projects/plus/SHOE/ OIL Ontology Inference Layer web site http://www.ontoknowledge.org/oil
Fens-2000 D. Fensel et al.: OIL in a nutshell In: Knowledge Acquisition, Modeling, and Management, Proceedings of the European Knowledge Acquisition Conference (EKAW-2000), R. Dieng et al. (eds.), Lecture Notes in Artificial Intelligence, LNAI, Springer-Verlag, October 2000.
DAML DARPA Agent Markup Language. http://www.daml.org