

# SEMANTIC WEB TECHNOLOGIES IN KNOWLEDGE MANAGEMENT

ASTA BÄCK, SARI VAINIKAINEN, CAJ SÖDERGÅRD AND HELENE JUHOLA

VTT Information Technology  
P.O.Box 12041  
FI-02044 VTT  
Finland  
tel. +358 9 456 1  
fax. +358 9 456 7052  
email: asta.back@vtt.fi, sari.vainikainen@vtt.fi,  
caj.sodergard@vtt.fi, helene.juhola@vtt.fi

Knowledge management is a big challenge especially in large organisations. Knowledge resides in many different forms: as explicit knowledge in documents and processes and as tacit knowledge in people and procedures and in many different forms between these two extremes. The vision of the Semantic Web is to offer more intelligent services by facilitating machine understanding of web content. Ontologies are an important building block in Semantic Web. An ontology describes the concepts, their relationships and properties within their domain, and it can be utilised both to offer automatic inferring and interoperability between applications. This is an appropriate vision for knowledge management, too. This paper describes how Semantic Web ontologies can be utilised in a research organisation to create a common language to describe its knowledge. The same ontology can be utilised to manage projects, people, documents and products. With a common ontology, information that is spread out in many different applications and documents can be viewable in a way that is easy to understand and navigate. The ontology makes it possible to search both knowledge content and experts who are linked to different topics, thereby bridging the gap between the tacit and explicit knowledge.

## INTRODUCTION

**Knowledge management** is a big challenge especially in large organisations. Knowledge resides in many different forms: as explicit knowledge in documents and processes and as tacit knowledge in people and procedures and in many different forms between these two extremes.

When an organisation wants to improve its knowledge management, it is important to focus is on strategically important knowledge. It is also important to clarify, how the organisation understands and defines knowledge. In [1], six different ways to define knowledge are listed. They are knowledge as personalised information, state of mind, object, process, access to information and capability. And there are even other definitions, such as knowledge as an asset. The best ways to manage knowledge depends naturally on which view on knowledge is taken, and then to choose the methods accordingly. There are many ways to view on the knowledge management process [2], the following five steps are often identified:

1. acquisition
2. creation,
3. storage,
4. validation, and
5. utilisation.

The vision of the **Semantic Web** is to offer more intelligent services by facilitating machine understanding of content. Ontologies are an important building block in the future Semantic Web. Ontologies provide a shared and common understanding of a domain that can be communicated across people and applications. This is an appropriate vision for knowledge management, too.

Originally, Ontology is a branch of philosophy that deals with the theory of being and considers questions about what is and what is not. In information technology, an ontology is the working model of entities and interactions in some particular domain of knowledge or practices, such as electronic commerce. In artificial intelligence (AI), an ontology is, according to Tom Gruber, an AI specialist at Stanford University, "the specification of conceptualizations, used to help programs and humans share knowledge." This seems to be one of the most, or even the most quoted definition for an ontology.

In this meaning, an ontology consists of specified concepts that are defined to create an agreed-upon vocabulary for information exchange. Knowledge in ontologies is mainly formalised using five kinds of components: classes, relations, functions, axioms and instances. [3]

Ontologies themselves can be classified into the following groups:

- Knowledge Representation ontologies
- General/Common ontologies
- Meta-ontologies, also called Generic Ontologies or Core Ontologies
- Domain ontologies
- Task ontologies

*Knowledge Representation ontologies* capture the representation primitives used to formalise knowledge in knowledge representation paradigms. The most representative example of this kind of ontologies is the Frame-Ontology, which captures the representation primitives (classes, instances, slots, facets, etc.) used in frame-based languages.

*General/Common ontologies* include vocabularies related to things, events, time, space, causality, behaviour, function, etc.

*Meta-ontologies*, also called Generic Ontologies or Core Ontologies are reusable across domains. A representative example of a meta-ontology may be a mereology ontology, which includes the term part-of.

*Domain ontologies* are reusable in a given domain. They provide vocabularies about the concepts within a domain and their relationships, about the activities that take place in that domain, and about the theories and elementary principles governing that domain.

*Task ontologies* provide a systematised vocabulary of the terms used to solve problems associated with tasks that may or may not be from the same domain. These ontologies provide a set of terms by means of which to generically describe how to solve one type of problems. They include generic names, generic verbs, generic adjectives and others in the scheduling tasks.

*Domain-Task ontologies* are task ontologies reusable in a given domain, but not across domains.

The two main *applications for ontologies* are conveying common understanding between applications (to create interoperability), and/or to automate some part of information processing. Next, some of the potential applications are discussed more in detail.

Probably the most important application is to make searches more effective. For example, a web site or a corporate intranet can organise its content according to some ontology, which then can be utilised to improve the quality of searches. For example, generalisation or specialisation of information can be utilised in assisting users. It is a short distance from general search applications to knowledge management applications. One of the big challenges in knowledge management is to find relevant knowledge and information, and here ontologies have a lot of potential. We can see application potential for both domain and task ontologies.

A closely related application to the above mentioned search applications is to support management of multimedia content. Ontologies can be used to provide semantic annotations for collections of images, audio, or other non-textual objects. These annotations can support both indexing and search. Since different people tend to describe non-textual objects in different ways, it is important that the search facilities go beyond simple keyword matching. Ideally, the ontologies would capture additional knowledge about the domain that can be used to improve retrieval of images.

One of the aims in the development of the Semantic Web is to make a shift from the tool paradigm to service paradigm. This means that now most applications only do the hard work, and humans are needed to make the actual decisions and control the application. With Semantic Web, the idea is to make more intelligent applications possible. For example, to schedule a meeting, the application could suggest the most suitable time according to specific user requirements and general user preferences. This requires ontologies that capture the concepts and knowledge in a deep enough level to make automatic applications possible. [4]

Another important area is ubiquitous computing, which includes different kinds of mobile applications. A key technology of true ad hoc networks is service discovery, functionality by which "services" (i.e., functions offered by various devices such as cell phones, printers, sensors, etc.) can be described, advertised, and discovered by others. All of the current service discovery and capability description mechanisms (e.g., Sun's JINI, Microsoft's UPnP) are based on ad hoc representation schemes and rely heavily on standardisation (i.e., on a priori identification of all those things one would want to communicate or discuss). The key issue and goal of Ubiquitous Computing is "serendipitous interoperability," and here ontologies play an important role. [5]

E-commerce applications from different domains is another important area, where ontologies are needed to enable interoperability at both the data and semantic levels. [Anon, 2003]

Also educational applications, in which students learn concepts and relationships directly from, or expressed in terms of, a common ontology, is a potential application area.

## **METHODOLOGY**

The research problem of our study was to assess how the emerging Semantic Web technologies can be utilised in knowledge and competence management within an expert organisation. This was studied with the help of constructive research. Constructive research is done by building a construction, e.g. a model or software application, to test its feasibility. The feasibility may be tested at two levels: can the construct be built, and if so, does it solve the problem.

The key idea was that the same ontology could be utilised to manage information about projects, people, as well as documents and products created in the projects by these people. This way it would be possible to keep the information up-to-date with minimum extra effort, and it would link different types of knowledge.

Our study defined knowledge primarily as access to knowledge that either resides in some document or in a person that has created the document or participated in a project that deals with certain topic or knowledge.

The idea of automatic updating, and links between different knowledge sources also address the knowledge storage issue.

## ONTOLOGY

We took ourselves, i.e. VTT, and particularly VTT Information technology, as the exemplary organisation where the ontology-based approach for an organisation's knowledge management was studied.

The following competency questions were defined for the ontology:

- Who knows about a certain topic?
- Which people are interested in a certain topic?
- What skills are being used in a certain project?
- Which projects deal with a certain topic?
- What is the competence of a certain research group?
- Which publications deal with a certain topic?

We chose to create our own ontology, but we utilised the work done with related ontologies. Our work was mostly influenced by the proper-ontology [6], which is an ontology for corporate skill knowledge, and KA2 ontology (Knowledge Annotation Initiative of the Knowledge Acquisition Community) [7], which is a large management initiative for the knowledge-acquisition research community. They both have been created by the University of Karlsruhe. New properties and resources had to be added to these ontologies to meet our specific requirements.

Our ontology consists of 11 main classes: KnowledgePortal, PortalTopic, ResearchTopic, Service, Skill, Profile, Project, Publication, Product, Organization and Person. These classes are enough to describe the research areas and to link the knowledge requirements and existing knowledge products to each other.

VTT knowledge portals were taken as the highest level of abstraction. These portals are the main way of presenting VTT's knowledge and skills to the outside world and represent internally a new connecting culture<sup>1</sup>. They offer clients an easy access to VTT knowledge and enable large and significant projects to reach ambitious goals. Portals nurture a culture to achieve innovations through producing, sharing and utilising tacit and explicit knowledge. Networking is the method to combine knowledge of VTT with the global innovative community. At present, VTT operates 8 knowledge portals, of which VTT ICT is the largest covering more than 800 people. VTT ICT has four knowledge areas: Enabling technologies, Telecommunication systems, Intelligent systems and services, and Tools for information society. Each of these areas is further partitioned into topics (see Figure 1).

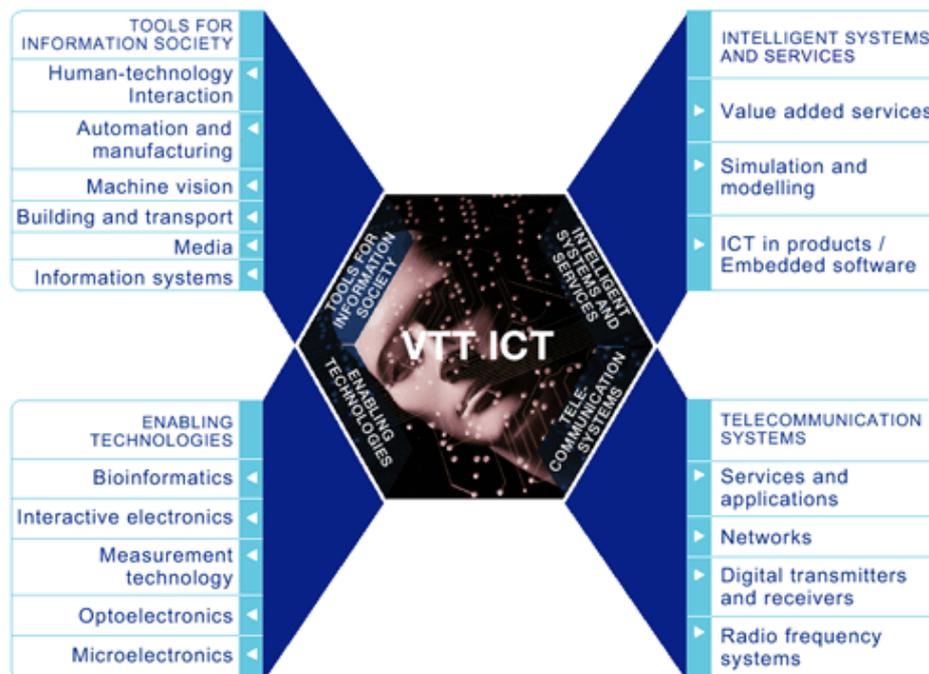


FIGURE 1. VTT'S ICT KNOWLEDGE PORTAL ([HTTP://WWW.VTT.FI/ICT](http://www.vtt.fi/ict)).

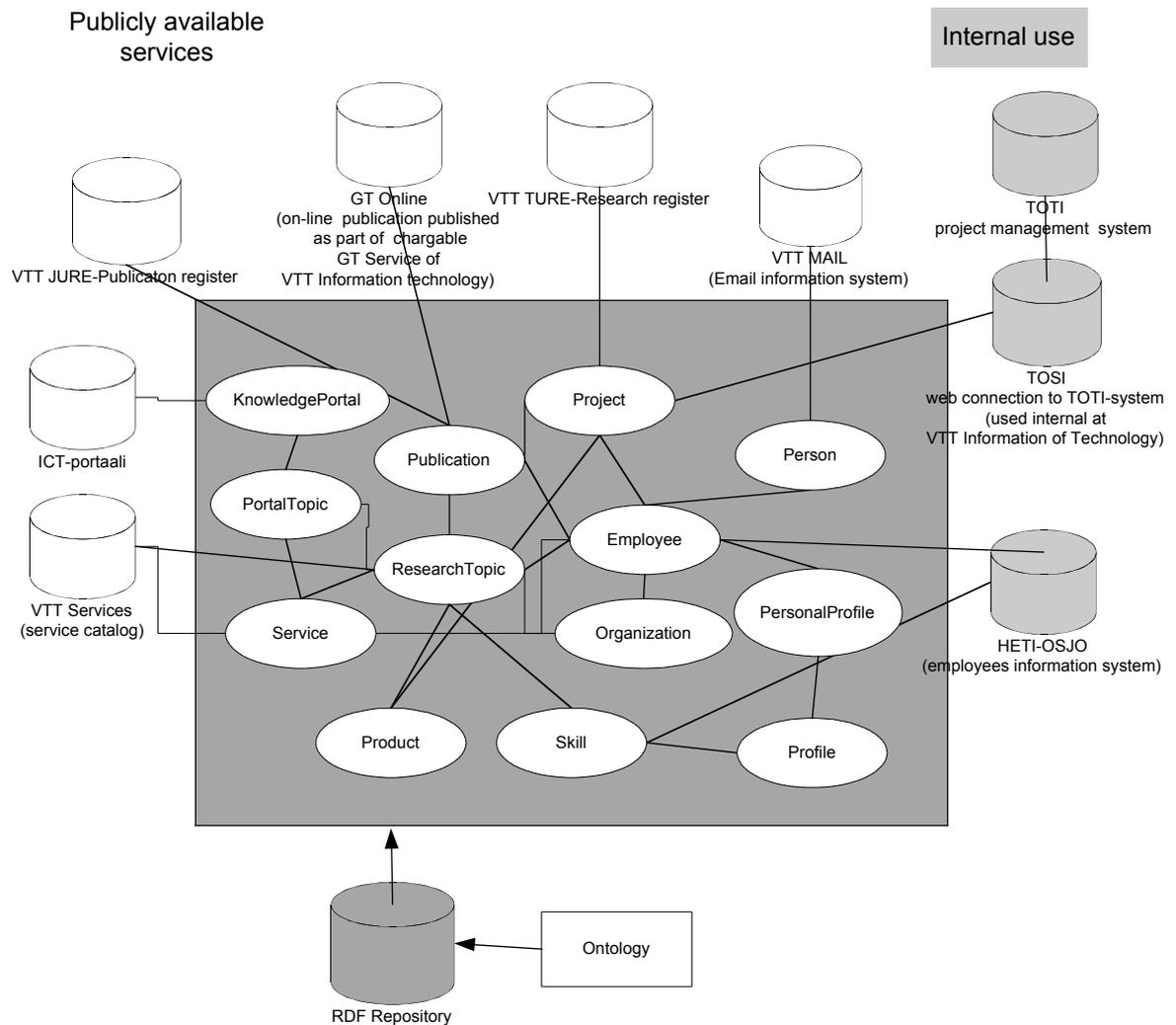
The idea was that information of people, projects or publications would be queried from existing databases, and that only new information is stored in the RDF repository to be built in this project.

<sup>1</sup> <http://www.vtt.fi/vtt/knowledgeportals/>

The most important existing knowledge sources are:

- the projects management system,
- the research project register,
- the publications register,
- an on-line publication,
- the employees information system,
- the service catalogue, and
- the email information system.

Links between the information in the existing knowledge sources and the ontology classes can be seen in Figure 2.

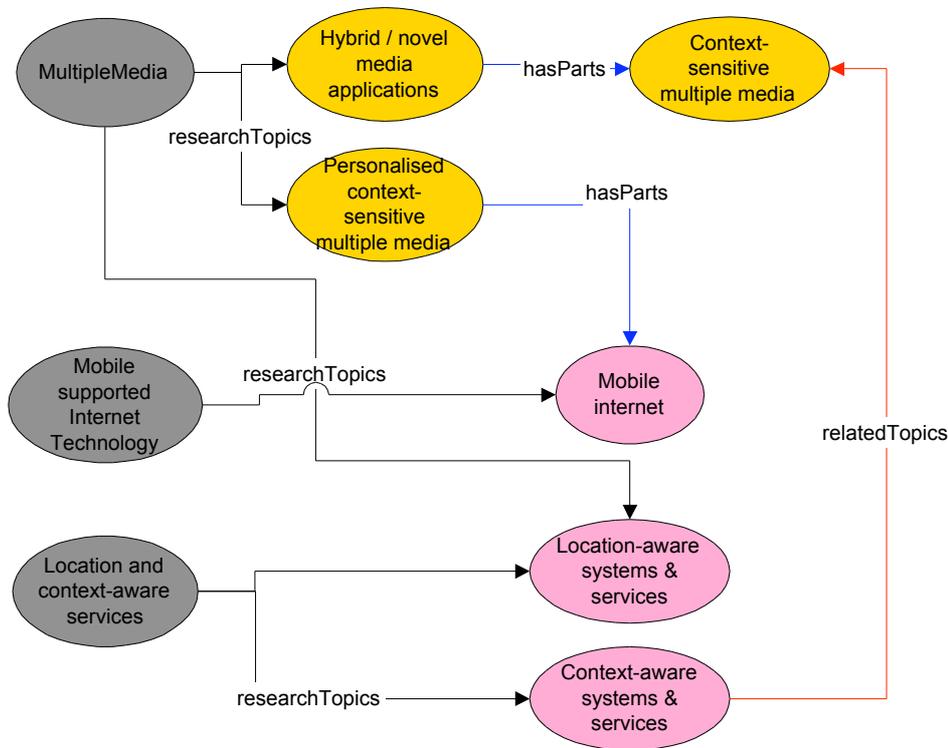


**FIGURE 2 – CONNECTIONS BETWEEN THE EXISTING KNOWLEDGE SOURCES AND THE ONTOLOGY RESOURCES.**

The KnowledgePortal class is connected to PortalTopics with the portalMainTopics –property. The PortalTopic class is connected to the ResearchTopic class utilising the researchTopics property. The difference between the PortalTopic and ResearchTopic classes is that ResearchTopics are assumed to be more permanent than PortalTopics. Research topics can be grouped in different ways and under different PortalTopic names depending on the strategy and focus. In similar manner, services are linked to PortalTopics with the help of the serviceTopics property. Both the research topics and services contain information of which portal they belong to; both these classes have the value for the belongsToPortal-property. The PortalTopic and ResearchTopic classes also have a relatedPortal property, which determines if the topic also relevant in some of the other VTT portals.

PortalTopic, ResearchTopic, Service and Skill all have a similar structure. All topics are determined as instances of the class. The instance hierarchy is created with properties. The hasParts property is used to describe which research topics are part of another research topic, and the relatedTopics property which ones are related

research to each other. The relatedTopics-property is needed to describe the fact that different research fields have close research topics, but they have a different view to it (See Figure 3). In similar manner, the Skill class has properties subSkills and relatedSkills.



**FIGURE 3 – AN EXAMPLE OF THE RELATIONS OF PORTAL TOPICS (GREY OVAL) AND RESEARCH TOPICS (OTHER OVALS).**

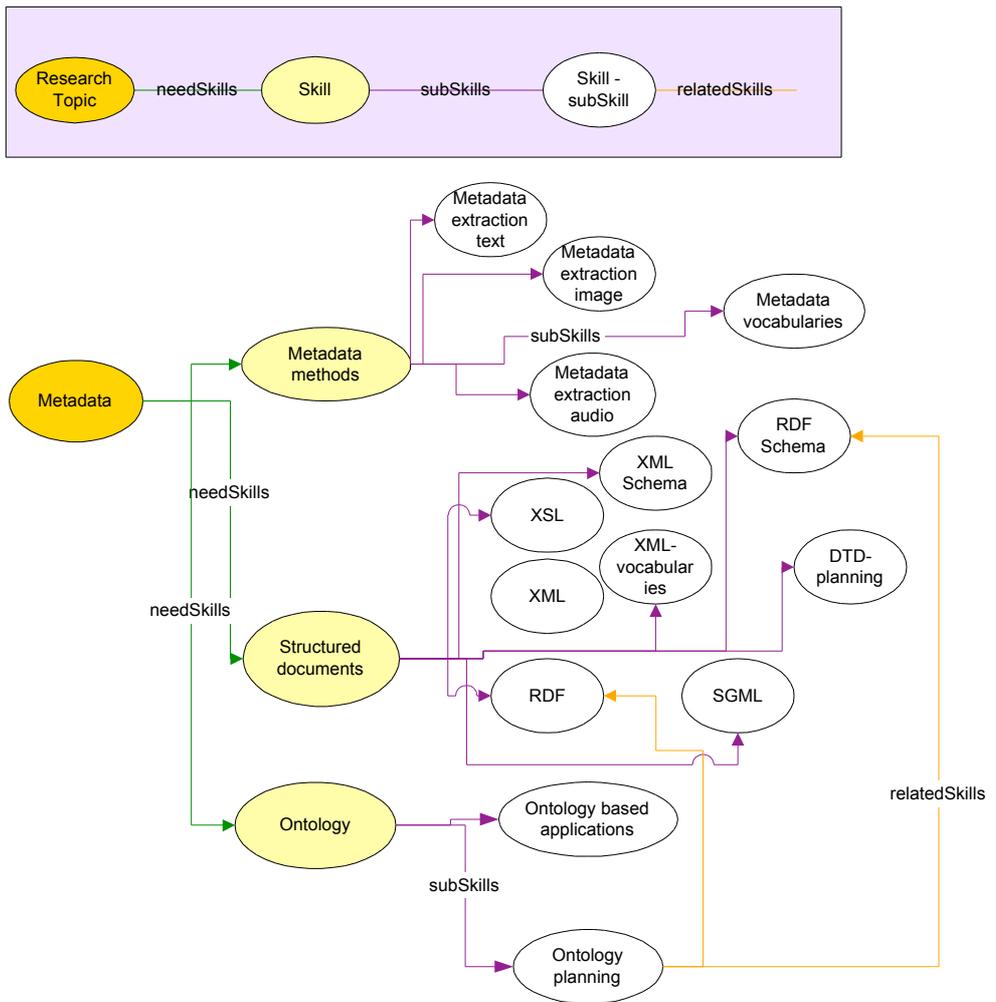
Research topics are connected to skills with a needSkill-property. As mentioned earlier, the skill hierarchy is determined with the subSkill and relatedSkill-properties. An example of the relations of a research topic and skills is shown in Figure 4.

The personalProfile-class (a sub class of Profile) determines an employee's personal competence profile. This profile is created with the evaluatedSkill-property, which determines the instance of skills to be evaluated. The level of competence is described with a value property. The skill competence level has four categories; no know-how, basic skills, has been applying at work and specialist. Projects can be connected to skills with the useSkills-property, which tells the skills that are being used in the project.

Employee (a sub class of Person) are linked to the projects they works for (worksAtProject-property), publications they have written (hasPublications-property), research topics they are interested in (researchInterest-property), and the research group they are a member of (memberOf-property). An employee's personal competence profile is connected to the employee through the belongToPerson-property in the PersonalProfile-class.

Projects are connected to appropriate research topics (isAboutTopic-property) and skills that they use (useSkills-property). Publications, products and services are connected to research topics.

The Organization class has three subclasses: ResearchInstitute, ResearchField and ResearchGroup. The structure of the organisation is described with the hasOrgParts property of the ResearchInstitute and ResearchField classes. Research topics are connected both to research groups (hasResearchGroupTopics property) and research fields (hasResearchFieldTopics property).



**FIGURE 4 – AN EXAMPLE OF THE RELATIONS BETWEEN THE RESEARCH TOPIC "METADATA" AND SKILLS RELATED TO IT (NEEDSKILLS-PROPERTY).**

The ontology contains quite a number of inverse-relations:

- worksAtProject property of Employee class and memberEmployee property of Project class
- hasProjectPublications property of Project class and relatedProject property of Publication class
- hasResearchPublications property of ResearchTopic class and isAboutResearchTopic of Publication class
- isAboutTopic property of Project class and hasProjects property of ResearchTopic class
- memberOf property of Employee and member property of ResearchGroup
- researchInterest property of Employee class and researchers property of ResearchTopic class
- hasResearchFieldTopics property of ResearchField class and researchFields property of ResearchTopics class
- has ResearchGroupTopics property of ResearchGroup class and researchGroups property of ResearchTopic class.

It is not necessary to store the information in both directions, because an application using the knowledge base can always infer the value of the inverse relation. However, from the knowledge-acquisition perspective it is convenient to have these both pieces of information explicitly available. When importing data into the system, it is enough to give the information in one place and the system can automatically fill in the value for the inverse relation. This way the consistency of the knowledge base can be maintained. Further ontology development will show if some of the inverse relations can be omitted.

## IMPLEMENTATION

A knowledge based was created describing one research area, and the idea was tested with the help of it.

### Search based on research topics

#### Select what you want to search

Persons  Projects  Publications  Products

#### Target your search

Search all that belongs to selected topic(s)

Target your search based on sub topics.

#### Select Research Topics you are interested in

##### Multiple Media

- Personalised context-sensitive multiple media
  - Metadata
  - Multiple channel and terminal publishing
  - Personalization
  - Information capturing and filtering
  - eLearning
  - Mobile Internet
- Community Media
  - Computer Games
  - Computer supported collaborative work
  - Knowledge management
- Broadcast Systems
- Hybrid / novel media applications
  - Hypermedia
  - Print to internet
  - Context-sensitive multiple media
- Integration and information systems in publishing processes
  - Information systems and processes in newspaper delivery
  - Intercompany information exchange
  - Process management in publishing and printing
- Location-aware systems & services

Next

**FIGURE 5 – A SEARCH BASED ON RESEARCH TOPICS. EXEMPLARY SELECTIONS ON THE FORM FOR FINDING THE PROJECTS, PERSONS AND PUBLICATIONS THAT RELATE TO RESEARCH TOPIC "KNOWLEDGE MANAGEMENT".**

An RDF Schema was created with the Protégé-2000 ontology editor, which has been developed at Stanford University. All the instances were saved as an RDF file. Both the RDF Schema and RDF data files were imported into the Sesame RDF-repository. Files were imported through the Sesame web interface. The Sesame repository was installed on top of the MySQL-database, where the case database was created. All the data was saved into database through Sesame. Queries into the knowledge base were evaluated utilising Sesame's RQL-editor. We did not implement the interfaces to other information systems.

The case implementation gives the users the possibility to make queries into the resulting knowledge base according to research topics or competencies. Persons and projects can be searched based on a research topics or competence. Publications and products can be queried based on research topics. Querying can be limited based on knowledge portal topics or a section of the organisation. An example of the search facilities based on research topics can be seen in Figure 5. Skill based searching skills has two options: direct search on skills (see Figure 6) and skills related to research topics. Standard queries, such as searching for certain employee's publications, could be implemented by making an interface to the already available systems that have this information.

## Search based on skills

### Select what you want to search

Persons  Projects

### Target your search

#### Select skills you are searching for:

##### Select expertise level:

all with some knowledge  no knowhow  basic skills  has been applying at work  specialist

##### Ontology

Ontology based applications  
Ontology planning

##### Metadata methods

Metadata extraction text  
Metadata extraction image  
Metadata extraction audio  
Metadata vocabularies

##### Publishing platforms

Digi-TV  
Print  
TV  
Mobile phone  
PDA  
Web

##### Personalization technologies

Collaborative filtering

##### Modelling methods

Context modelling  
User modelling  
Product modelling

##### Databases

SQL-databases  
XML-databases  
RDF-databases  
Ontology databases  
Multimedia databases  
Database planning and implementation  
Database installation and maintenance

##### Structured documents

XML vocabularies  
DTD planning  
XML Schema  
RDF Schema  
SGML  
XML  
RDF

**FIGURE 6 – SEARCH BASED ON SKILLS. SEARCHED SKILLS ARE SELECTED FROM THE LIST. EITHER PERSONS OR PROJECTS CAN BE SEARCHED BASED ON SKILLS.**

## DISCUSSION

This study addressed the issue of combining information from different sources, where an ontology plays the role of tying them together in order to assist people in finding relevant information.

When designing the system further, several aspects must be taken into consideration. Securing the data and respecting the privacy of the users are important, even though somewhat at odds with the main goal of increasing knowledge sharing within the organisation. It would also be useful to make a visual version of the ontology to help the users to understand more easily the connections between domains and the available information.

Gathering information of user actions, and utilising this information to update profiles, is another option for further development. Another application potential for the ontology could be to use it as a tool to evaluate what the organisation knows as a whole, and identify organisation's knowledge strengths and gaps, also with relation to future demands, i.e. to assist in strategy management.

In a large organisation like VTT with more than 3000 people and six research departments, defining all the ontology instances and their relations is a considerable task. The work can and actually must be shared between the different research groups so that each of them is responsible for describing their own specific field. When doing this work, it is important that all participants have a clear focus on what the ontology is expected to accomplish.

It would also be good to have a common understanding of what is the level of detail that should be aimed at, even though an ontology supports managing information of research topics and skills at varying levels

of detail. For example, one may wonder if it is enough to determine ontology planning as a skill or should we go deeper and determine the competence in the different ontology languages like RDF, OIL and DAML. We chose not to go very much into detail, because our assumption was that in most cases it is enough to know that someone is a specialist in some upper level topic like ontology planning. If someone is searching for an expert to solve a particular detailed problem or looking for an expert for a project, it is most efficient to contact the person, and ask for more information.

We also noticed that the difference between a research topic and a skill is not always clear. For example, a research field may have databases as their research topic, and they actually develop database technologies, whereas there are other research fields that also have database related competencies, but only to utilise databases in their projects.

## CONCLUSIONS

Offering and finding relevant information and knowledge is a huge challenge. It is important in the private lives of people, and even more important in companies and organisations, which are faced with even increasing competition. It is said that knowledge and the ability to learn quickly are the only permanent competitive advantages that a company may have. Semantic Web technologies are a tool that can be used to better manage information and to increase the level of automation in knowledge and information acquisition tasks.

In the eighties, a lot of expectations were focused on developing artificial intelligence technologies. They, however, failed in many ways. Semantic web technologies try to address the same challenges, but with a more modular and a more light-way approach. Ontologies are the key element in semantic web technologies. The web has made it possible to easily share ontologies, and work is under way to develop ontology languages to improve the management of distributed ontology vocabularies.

Knowledge maps and technology roadmaps have become a popular way of capturing knowledge about people's know how and visualising companies' technological positions and future opportunities and threats. A knowledge or technology map can be expressed as an ontology. This was done in our case application. When the ontology is used as a consistent access point to anything that an organisation does, it can assist in guiding users to relevant knowledge be it in tacit or explicit format.

It should be possible to import information of new projects, documents and people into the system automatically utilising existing systems and management processes, so that manual updating and inputting new information can be minimised. The ontology gives us the opportunity to do that. For example, when a project belongs to certain research topic, we can infer that also the publications of the project relate to same topic, or if an employee works in a project, that he or she has skills relevant in that project. The strength of the ontology based approach is that we can utilise information from different points of view, thereby creating opportunities for automatic updating different kinds of information. The ontology-based approach also bridges the gap between tacit and explicit knowledge, because it supports managing information of documents and people within the same system.

The tools for ontology development, as well as the ontology languages are still under development. This means that there are many research challenges and opportunities, and that it is early to build real industrial applications. The aim in developing future knowledge management and personalisation application could be to increase the level of automation on both providing relevant knowledge and information sources, and also in analysing and updating information. It is difficult to automate these tasks completely, but with increased level of automation, it is possible to improve the user experience and increase the productivity of knowledge work.

We could describe an ontology meeting our requirements utilising RDF. There is, however, little support for ontology management, like for different version, which would be important in a real-life application.

Creating an ontology for a research institute requires determining the instances of research topics and their related skills, and it is a demanding task. Nevertheless, already the development phase is very useful, because it requires analysing the research area and the necessary fields and levels of expertise. Large research organisations have people with same research interests and expertise in different locations, in different research fields and even in different research institutes. The ontology development helps in revealing potential for co-operation beyond the organisational limits.

It is also important to remember that the ontology and the knowledge base built on it needs to be updated and even modified as things change. The tools should support this, because otherwise the system quickly becomes outdated.

The work supported knowledge management by showing how to give better access to information and knowledge. This can be regarded only as the first but as a very necessary step in improving knowledge management. When access to new knowledge and links to other people with complementary knowledge are found, there is an opportunity to go forward to the next steps, knowledge creation, validation and utilisation.

1. ALAVI MARYAM, AND LEIDNER, DOROTHY: Review: Knowledge management and knowledge management systems: conceptual foundations and research issues. MIS Quarterly Vol 25 No 1, pp. 107-136 (March 2001)
2. RUBENSTEIN-MONTANO B., J. LIEBOWITZ, J. BUCHWALTER, D. MCCAW, B. NEWMAN, K. REBECK. A systems thinking framework for knowledge management. Decision Support Systems 31(2001)5-16. Elsevier.
3. OSCAR CORCHO, MARIANO FERNÁNDEZ-LÓPEZ, ASUNCIÓN GÓMEZ PÉREZ. Technical Roadmap v1.0. 30/11/01. OntoWeb research project document. 74 p. <http://ontoweb.aifb.uni-karlsruhe.de/About/Deliverables/Deliverable111.pdf>
4. LASSILA, ORA. Towards the Semantic Web. Keynote address at XML Finland 2002 Conference. October 21-22, 2002. Helsinki. Finland.
5. SEAN BECHHOFFER, CAROLE GLOBE, IAN HORROCKS. Requirements of Ontology Languages. IST Project. IST-2000-29243 Ontoweb. Deliverable 4.1. 22.3.2002. 19 p. <http://ontoweb.aifb.uni-karlsruhe.de/About/Deliverables/d4.1.pdf>
6. YORK SURE, ALEXANDER MAEDCHE AND STAFFAN STAAB (University of Karlsruhe). Leveraging Corporate Skill Knowledge - From ProPer to OntoProPer. Proc. of the Third Int. Conf. on Practical Aspects of Knowledge Management (PAKM2000). Basel, Switzerland, 30-31, Oct. 2000. 9 p. [http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-34/sure\\_et\\_al.pdf](http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-34/sure_et_al.pdf). (See also The proper-ontology. <http://ontobroker.semanticweb.org/ontologies/proper-onto-2000-11-07.flo>)
7. ANON. KA2-Ontology <http://www.ontoknowledge.org/oil/case-studies/> (See also KA2 Knowledge Annotation Initiative of the Knowledge Acquisition Community–ontology (<http://ontobroker.semanticweb.org/ontologies/ka2-onto-2000-11-07.flo>))