

Semantic Web Based Integration of Knowledge Resources for Supporting Collaboration

Vili Podgorelec, Luka Pavlič and Marjan Heričko
 Institute of Informatics, University of Maribor, FERI,
 Smetanova ulica 17, SI-2000 Maribor, Slovenia
 E-mail: vili.podgorelec@uni-mb.si

Keywords: knowledge management, semantic web, collaboration, knowledge sharing

Received: October 27, 2006

In the paper we present the importance of collaboration between researchers for the improvement of their creativity. A unified methodology to support collaboration strategies of researchers and research teams based on knowledge sharing is introduced. We argue that a defined methodology together with an efficient technological system supporting the methodology should improve the creativity of research teams and consequentially facilitate the development.

Based on the required functionalities of such a system, we propose the semantic web as the underlying technology. It is indicated how the semantic web technologies could provide the necessary solutions for the integration of data resources, the transformation of data into valuable knowledge, the effective use of knowledge by intelligent information services and knowledge sharing both within an organization and inter-organizationally. Finally, the prototype architecture of an intelligent agent within a database system is outlined, which serves as an information integration mechanism.

Povzetek: Semantična integracija virov znanja v podporo sodelovanju.

1 Introduction

Creativity is a complex cognitive activity for the performing of which both motivation and knowledge are required. A motivation is partially provided by a working environment in which a researcher or a research team performs its activities. The most important aspects for a high motivation are:

- effective fulfillment of conditions required for creativity, in which the researchers are able to optimally put into practice their potentials, and
- efficient support system that enables researchers to solve their problems and/or overcome obstacles, which they may experience during the creativity process.

Naturally, for the successful solving of problems the second factor of creativity process is of vital importance, namely knowledge. If the researchers are supposed to be creative, they need to possess the knowledge that will enable the creativity. Many organizations performed various studies which confirmed that their research staff is professionally highly skilled, however, their creative results were not comparable with the leading teams. The reason could be an inadequate or an ineffective approach to collaboration between single researchers and/or research teams when performing more complex research projects.

It is hard to believe and yet true that researchers within an organization usually do not know the knowledge and skill profiles of their peer researchers.

Consequentially, the problem occurs when individual researchers, who possess a good amount of individual knowledge, face a problem they are not able to solve by themselves and that could be effectively solved by some of their peers. Many a time a solution is not reached because of an inadequate or even non-existing knowledge sharing. On the other hand, the researchers who can not put their knowledge and skills into practice become less and less motivated.

In both situations the consequence is a lower creativeness and in the worst cases even the total suppression of the creative energy. And without creativity there can not be any real development. Knowledge sharing and collaboration is deficient within organizations, even within research departments and institutes. On the inter-organizational level, there is almost the total lack of systematically organized and planned collaboration. Regarding this, the definition and development of a proper methodology supporting collaboration of researchers, effectively initiated into research and development departments, could importantly contribute towards higher creativity and consequentially to faster and more efficient development.

2 Knowledge based collaboration between researchers

Based on the awareness of how important creativity is for the efficient development and the vital part of knowledge

within this process, a lot of researchers studied different aspects of knowledge management and assets management systems for knowledge capturing, representation and sharing [Art06]. Various theories on exploitation of knowledge sources within organizations and workgroups are being introduced by scientists; however, they are not technically and technologically supported. On the other hand many companies decide to set up a knowledge management system that is not efficiently used to their advantage, because of the lack of adequate methodology.

From the technological point of view, the necessity of transition from software products to services has been globally recognized. In this manner, there are many attempts to set up a knowledge-based system based on ontologies and semantically annotated data. Nevertheless, usually the ontologies are used primarily for statically describe data repositories. In the very near future we predict the intelligent approaches to more complex information services which will need to make advantage of semantically annotated repositories.

The literature search showed that presently there is no completely defined and technologically supported methodology to support researchers' collaboration strategies based on knowledge. However, a set of single approaches and solutions gives evidence of a possibility to define and develop such a methodology. In our opinion, there are several attempts which individually provide basic components for further development.

An interesting approach to bridging communities of practice with information technology in pursuit of global knowledge sharing is presented in [Pan03]. A similar approach to knowledge sharing in an emerging network of practice is presented in [Baa05]. They both suggest a use of knowledge portals, which are an extension of well known and recently much used business portals for managing all important business data. An interesting framework for stimulating innovation is presented in [Bre05] that gives evidence for the importance of properly technologically supported methodology of collaboration to improve the creativity. The importance of collaboration based on knowledge is recognized also in [Lom06], where the authors suggest a framework to manage formalized exchanges during collaborative design. The inter-organizational resource sharing decisions in collaborative knowledge creation is especially emphasized in [Sam06]. We also proposed a possible solution to build project teams based on knowledge with the use of information technology [Pod06].

3 Methodology to support collaboration of researchers

The research problem that needs to be addressed is the definition and development of a proper methodology to support strategies of collaboration between researchers based on knowledge management and assets management. Such a methodology should contribute to a more optimal access to knowledge and competences within a research area. Furthermore, it should improve

the creativity of researchers and lead towards more efficient development, both from the organizational and the technological point of view.

3.1 Outline of collaboration methodology based on knowledge sharing

We plan to achieve the proposed methodology by using one of the most vibrant of today's information technologies, namely semantic web. In our opinion the semantic web is a proper technology to bridge the technological gaps, outlined in the present approaches to a unified knowledge-based collaboration methodology.

The most important aspect of semantic web technologies is the semantics of data, which allow efficient integration of information resources (both existing and forthcoming) and a possibility of automatic, intelligent inferring on knowledge, retrieved from those information resources.

Having the technology for the efficient integration of information resources and the technology for semantically annotate these data, a step to transform the data into knowledge becomes possible. Having a unified access to knowledge resources of an organization (i.e. a logical organizational unit, like a department) an information service that uses this knowledge becomes reasonable. Having the information services which use the knowledge resources of an organization to provide the useful functionalities a system to share knowledge within an organization and also inter-organizationally exceeds mere theories and hoped-for ideas. The schematic outline of the proposed collaboration methodology based on knowledge sharing is presented on Figure 1.

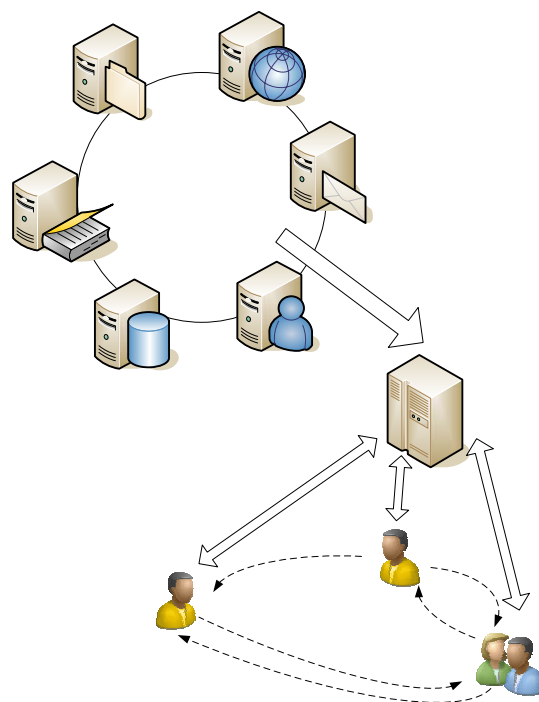


Figure 1: Collaboration based on knowledge sharing.

3.2 Managing individual competences to improve organizational development

Additionally to sharing organizational knowledge in order to improve the overall creativity and development within an organization, it is also important to manage individual knowledge competences in order to establish a knowledge map of an organization [Col03]. In this case a researcher or a research group within an organization is able to locate those who possess the required expertise for solving a specific task or performing a special activity. In this way not only the creativity could be considerably increased, but also the knowledge and skills of individual researchers are improved, because they learn at first hand from colleagues which master a specific issue.

On the other hand, when performing research projects, a project leader or team members can easily recognize which colleagues are appropriate for performing different project tasks. Also the hidden skills, not directly stated in a profile of an individual researcher, can be discovered by a proper skills matching approach. Again, this is possible by having all the data integrated and semantically annotated, what allows the support system to automatically infer on the stored data. A part of the knowledge portal for managing individual profiles is presented on Figure 2.

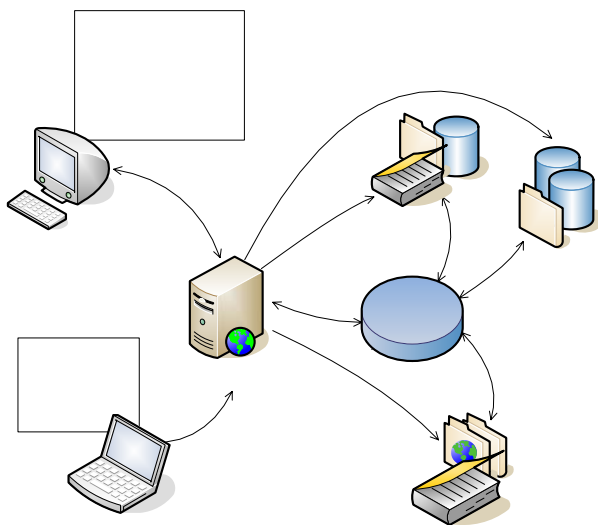


Figure 2: Semantic knowledge portal for managing personal skills profiles.

4 Semantic web as the underlying technology

In order to fulfill the requirements, which are necessary to achieve the proposed methodology, semantic web technologies are in our opinion a very sound choice. The semantic web represents one of the most vibrant of today's information technologies. As it turns out, it is very appropriate for the integration of information

resources by semantic annotation of data. Furthermore, the idea of semantic web allows automatic, intelligent inferring on knowledge, retrieved from these information resources.

4.1 Semantic web technologies

The basic idea of semantic web is a different organization and storage of data and consequentially new possibilities to use this data [Ber01]. Although the idea of semantic web is based on well established concepts, such as machine learning and automatic reasoning, the semantic web community has given these areas fresh new move by introducing web-based solutions. The web is still very primitive and actually provides quite useless organization of knowledge, especially when one want to do some searching or knowledge discovery. The barrier that prevents more advanced usage of the web is believed to be semantic poorness of today's world wide web. Data, documents, images and all other kinds of content on the web are presented as very simple, non structured human readable and human understandable materials. The result is the inability to make a real use of the web's enormous amount of "knowledge". Because it can be understood as a huge cross-referenced library, all we have is by default a weak tool called keyword search.

In order to overcome those difficulties the concept of meta-data is introduced on the web. Using meta-data, so called smart agents can be used for searching by content. As a foundation, there has been a lot of work done about common formats for interchange of data and common understanding of common concepts. That allows a person or a machine to browse, understand and use knowledge on the web in a more straightforward way. All those activities and technologies are known by the term "semantic web".

Furthermore, the semantic web ideas and technologies can be used in other areas also, not only on globally available web. They can be used in the enterprise information systems for knowledge management in a different way to introduce new intelligent services.

As already mentioned, we want knowledge (with its meaning!) to be accessible to both people and machines. It is obvious that we need to represent knowledge in a more formal way. There are quite a lot of possibilities. The most appropriate for semantic web were chosen semantic nets. They are very simple nets, consisting of linked concepts. The question is what we need to represent distributed knowledge, such as we have on the web? We need a standardized way of naming things. Two different things should have different names and vice versa, when we talk about the same thing we need to use the same name. Furthermore, we need a standardized way of saying something about things – we need a standardized way of describing things. Also we need common vocabularies. If we talk about coin and bank note, for example, we should automatically know that we are talking about money. And finally, we also need a standardized way of giving semantics to data, or said

more technically, we need a standardized technology to connect data with some meta-data.

In semantic web, knowledge is represented as nets, written down in XML-based language called RDF (Resource Description Framework). RDF is dealing with URIs (another W3C standard for naming resources globally unique). Advanced use of semantically annotated data can only be accomplished by using ontologies represented as a RDFS or OWL documents. The whole stack of technologies for semantic web is shown on Figure 3.

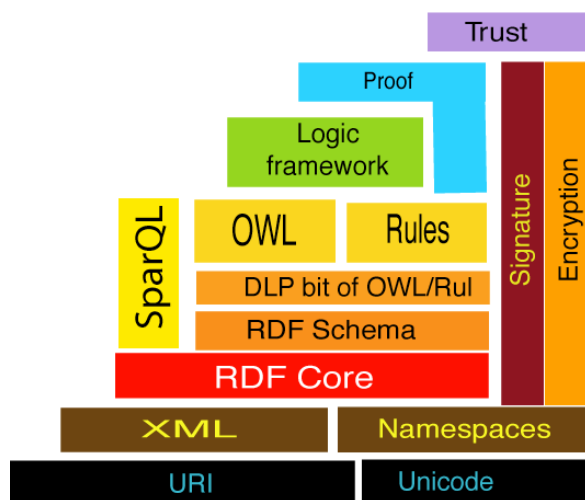


Figure 3: The complete stack of semantic web technologies as proposed by Tim Berners-Lee and W3C.

4.2 The key role of ontology in knowledge management

What is the purpose of ontology in semantic web? Ontology describes the subject domain using notions of concepts, instances, attributes, relations and axioms. In [Gru93] authors define ontology as a formal explicit specification of a shared conceptualization. It is a useful way to organize and share information while offering an intelligent means for knowledge management. Ontology also enhances semantic search in distributed and heterogeneous information services. Ontologies are the key player, if we want to do (automatic) search in more advanced ways, not only keyword search.

There are several benefits of using ontologies for information solutions. Semantic search engines return instances that constitute answers to queries rather than documents containing search strings as in keyword search engines. Semantic search uses meanings (semantics) of the query terms defined in the ontology. The data of ontology constitutes precise answers to user questions. Users can further browse related concept because answers are interconnected through semantics. It can be speculated that using ontology supported systems users will also be able to invoke functionalities or query data using free text input in the future.

4.2.1 The definition of domain knowledge and personal skills ontology

In order to adopt the semantic web technologies in our pursuit of implementing the proposed methodology, there are two key fields which need to be addressed: domain knowledge that we want to share between researchers and the personal skills profiles of the researchers. For those two areas the ontologies need to be defined, which will then allow all further actions, like semantic annotation of data (in accordance with the ontology), integration of data resources, advanced searching and inferring on the data.

The definition of domain knowledge is closely inter-related with the definition of personal skills for this domain. For example: let's imagine the field of software engineering where we want to describe the programming skills of researchers. If we want to adequately define the technical skills of a Java programmer, it is important to know at least the basic attributes of Java programming language. On the other hand, if we want to describe a "knowledge item" such as Java source code, it is very useful to link it with the knowledge requirements of producing such an item. Only in this manner knowledge sharing (intelligent searching, inferring, etc) could be successful.

Naturally, the definition of domain knowledge for a specific area may vary, but there have been some attempts to define personal skills ontology. One of the most important features of semantic web is the fact, that the defined ontology can be easily broadened when a need occurs for a new concept, not previously identified. In this manner, a working system can be expanded with new knowledge on the fly.

4.2.2 Ontology-based personal skills management

An overview of the related work in ontology-based personal skill management is presented in [Bie05]. Already [Sta99; Jar99] promoted the idea of ontology-based modeling of personnel skills and job requirements – as part of comprehensive, workflow-oriented enterprise modeling. There, the following potential applications of ontology-based skill profiles are listed:

- skill gap analysis – at the enterprise level, as a part of strategic HR planning,
- project team building,
- recruitment planning – again a part of strategic HR planning, and
- training analysis – at the level of individual personnel development.

Those approaches were mainly technology-driven and were – to our knowledge – never realized in a large-scale industrial environment. Nor have they been accepted by the HRM departments, translated into HRM people's terminology, embedded into more comprehensive models and procedures of HRM people, and integrated with existing software infrastructures.

After those first publications, there were a number of interesting technology-oriented researches which showed that in particular skill matching can benefit from interesting technological approaches, such as background knowledge exploitation. For instance, [Liao99] employs declarative retrieval heuristics for traversing ontology structures. [Sure00] derives competency statements through F-Logic reasoning and developed a soft matching approach for skill profile matching. [Colucci03] and others use description-logics (DL) inferences to take into account background knowledge as well as incomplete knowledge when matching profiles.

4.3 Integration of data resources

As we believe in the applicability of semantic web technologies for knowledge sharing, ontologies and semantically annotated data are used for describing personal skill profiles and domain knowledge. In this manner the existing knowledge within organizations can be reused, if it is appropriately converted into RDF in accordance with the defined ontology.

The proposed methodology envisages an integration of many data resources, containing the necessary knowledge. In this manner the integrated information resources can be seen as an inter-connected database. To allow an easy access to integrated knowledge resources, our technical solution enables one to use well-known queries, like SQL or XQuery, to access this integrated data universally (Figure 4).

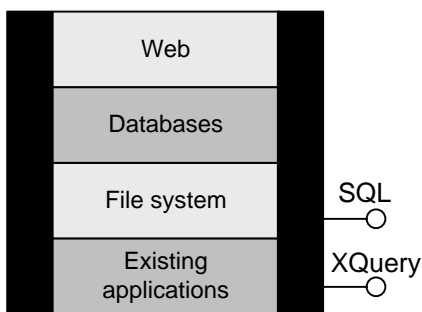


Figure 4: Universal access to integrated data resources.

4.3.1 The architecture of the system

The architecture of the system has been designed in a form of four inter-connected main components (Figure 5). RDF Provider is in the role of an intelligent agent that continuously acquires data from all available data resources. The collected RDFs are stored within a XML database, where they are used by the user interface component. This component is used for storing data, browsing, searching and inferring on the semantically annotated data. The communication with XML database is realized using standardized interfaces, like SQL and XQuery.

The stack of used technologies is presented on figure 6. The system prototype is implemented mainly in Java

programming language using open source Jena semantic web development library [JENA]. It provides us with a straight-forward development system, very appropriate for semantic web portal application. Because the inferring technology, as represented in the stack of semantic web technologies proposed by W3c (Figure 3) is not available yet, for the inferring part of the system we chose CLIPS programming environment [CLIPS]. It is a production rule based programming system mainly used for developing expert systems. CLIPS is a productive development and delivery expert system tool which provides a complete environment for the construction of rule and/or object based expert systems. As it turned out, it enabled us with powerful inferring possibilities. Additionally, it is very easy to execute CLIPS rules within Java applications using open source libraries such as JClips [JCLIPS], what further enables one to integrate the inferring rules within an information system. The whole system itself does not consist of many different technologies, which is in our belief good. The fundamentals of the system lies in J2EE platform and XML enabled database.

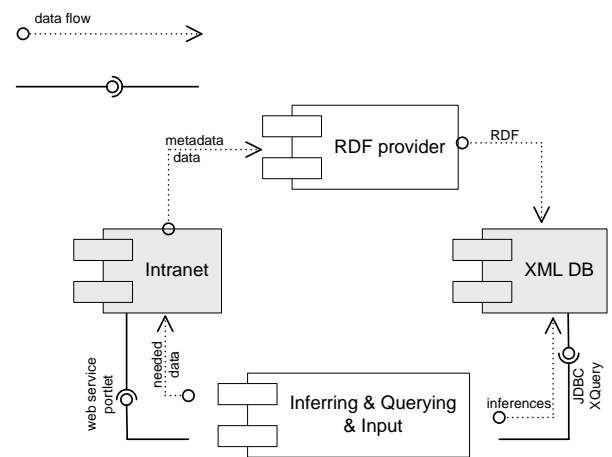


Figure 5: Main components of the systems and their inter-connections.

The most important component, called RDF provider, is responsible for collecting as many internal data in RDF as possible. As it continuously examines all possible data providers within the specified range and domain, it extracts data from existing applications, web pages, databases and file systems. Collected RDFs and presented ontology are persistently stored in XML database and prepared for analysis with reasoning system, based on J2EE, Jena and CLIPS.

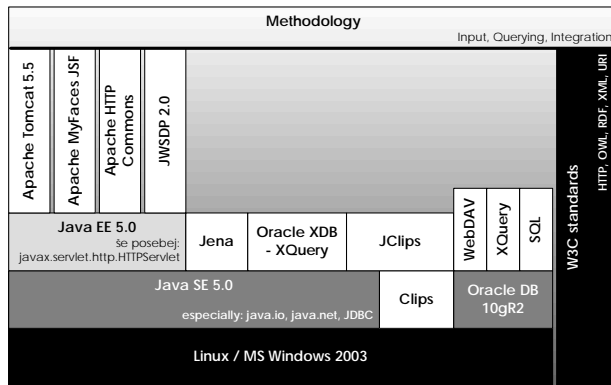


Figure 6: The stack of used technologies.

5 Issues to be resolved

In order to achieve the proposed methodology, a number of current challenges and aspects need to be addressed. Semantic web technologies as an underlying technological framework represent a vibrant new technology with high potential, and yet as a complex approach require several scientific and technological solutions. The semantic web potentials as a technique for integrating existing and forthcoming information solutions with semantics need to be addressed. Also the approaches to automatic annotation of data and the intelligent web services (like automatic discovery of hidden knowledge, automatic profiles matching, project teams building) are an issue. Because the intention is to provide efficient bridging of research communities with information technology based on knowledge, the possibilities of automatic construction of knowledge from data need to be studied (in combination with existing powerful approaches like data mining, text and web mining, knowledge discovery from data), as well as the linkage of ontologies and semantic repositories.

Finally, as of our knowledge, the challenge of reducing complexity by systematic linkage of research groups has not been adequately answered yet. It is our belief that the defined methodology can contribute a great deal to answering also this important question.

6 Conclusion

Our view of a unified methodology to support collaboration strategies of researchers and research teams based on knowledge has been presented in the paper. In order to improve the creativity of research teams and consequentially also facilitate the development, a defined methodology together with an efficient technological system supporting the methodology could be the right way.

Based on the required functionalities of such a system, the semantic web can be used as the underlying technology. It has been indicated how the semantic web technologies can provide the necessary solutions to the integration of data resources, the transformation of data into valuable knowledge, the effective use of knowledge by intelligent information services and knowledge

sharing both within an organization and inter-organizationally.

When a proposed methodology is developed, it could be used to semantically describe the competence profiles of researchers, involved in research groups of various organizations. In this way a considerably better collaboration of researchers would be achieved within a scope of scientific, research and development activities. Furthermore, the researchers from academic institutions and industry could be efficiently inter-connected, what would in turn lead towards higher creativity and faster industrial development. In fact a lot of data, needed for the operation of such a system, is already available and stored within different databases (researchers' profiles, publications and research activities, project data and project teams data, description of research projects and their results, ...) All this data only needs to be appropriately annotated and integrated, which is an inherent property of the proposed methodology. The existing information services, although not semantically annotated, could be used in efficiently integrated within the system by implementing the required interface wrappers.

References

- [Art06] H.A. Artail, Application of KM measures to the impact of a specialized groupware system on corporate productivity and operations, *Information & Management*, 43(4), 2006
- [Baa05] P. van Baalen et al., Knowledge Sharing in an Emerging Network of Practice:: The Role of a Knowledge Portal, *European Management Journal*, 23(3), 2005
- [Ber01] T. Berners-Lee, Business Model for the Semantic Web, <http://www.w3.org/DesignIssues/Overview.html>, 2001.
- [Bie05] E. Biesalski, A. Abecker, Integrated Processes and Tools for Personnel Development, *Proc. of 11th International Conference on Concurrent Enterprising*, Munich, Germany, June 2005
- [Bre05] A. Brennan, L. Dooley, Networked creativity: a structured management framework for stimulating innovation, *Technovation*, 25(12), 2005
- [Col03] S. Colucci, T. Di Noia, E. Di Sciascio, F.M. Donini, M. Mongiello, M. Mottola, A Formal Approach to Ontology-Based Semantic Match of Skills Descriptions, *Journal of Universal Computer Science*, Springer Verlag, 9(12), pp. 1437-1454, 2003.
- [Gru93] T.R. Gruber, Towards Principles for the Design of Ontologies used for Knowledge Sharing, In N. Guarino & R. Poli (eds.), *Proc. of International Workshop on Formal Ontology*, Padova, Italy, 1993
- [Jar99] P. Jarvis, J. Stader, A. Macintosh, J. Moore, P. Chung, What Right Do You Have to Do That?, *Proc. of ICEIS - 1st Int. Conf. on*

- Enterprise Information Systems*, Portugal, 1999
- [Lom06] M. Lombard, L.G. Yesilbas, Towards a framework to manage formalised exchanges during collaborative design, *Mathematics and Computers in Simulation*, 70(5-6), 2006
- [Pan03] S.L. Pan, D.E. Leidner, Bridging communities of practice with information technology in pursuit of global knowledge sharing, *The Journal of Strategic Information Systems*, 12(1), 2003
- [Pod06] V. Podgorelec, L. Pavlič, M. Heričko, Using semantic web technologies for project team building, *Proc. of International Conference on Knowledge Management in Organizations KMO-2006*, June 2006
- [Sam06] S. Samaddar, S.S. Kadiyala, An analysis of interorganizational resource sharing decisions in collaborative knowledge creation, *European Journal of Operational Research*, 170(1), 2006
- [Sta99] J. Stader, A. Macintosh, Capability Modelling and Knowledge Management, *Applications and Innovations in Expert Systems VII*, Springer-Verlag, pp. 33–50, 1999.
- [CLIPS] –, CLIPS – A Tool for Building Expert Systems, <http://www.ghg.net/clips/CLIPS.html>
- [JCLIPS] –, JClips — CLIPS for Java, <http://www.cs.vu.nl/~mrmenken/jclips/>
- [JENA] –, Jena – A Semantic Web Framework for Java, <http://jena.sourceforge.net/>