

A Semantic Web Portal to construction knowledge exchange

M. Argüello, A. El-Hasia & M. Lees

Research Institute for Built and Human Environment (BUHU), UK

Abstract

A current construction challenge is the creation of a construction university to help promote innovation and best practice throughout the industry, with the aim of developing a strong network of organisations able to share knowledge at different levels. Establishing the Construction Knowledge Exchange will be an important first step towards this aim. The Semantic Web provides an appropriate platform for knowledge exchange, and the benefits of implementing Semantic Web technologies can be easily identified as Semantic Web technologies have the potential to increase the information consistency and the information processing quality of Web portals. To exchange knowledge, i.e. semantic exchange, requires a shared mechanism to classify domain knowledge-items or information into interrelated concepts, i.e. ontology. The ontology has been developed in the OWL Web Ontology Language and is the backbone of a Semantic Web Portal (the β -CKE portal). The β -CKE portal provides appropriate usability and availability to represent a user-friendly interface for a virtual community, and furthermore advanced functionalities for community users where enhanced search accomplishes significantly better search results than other information retrieval techniques. The evaluation of the β -CKE portal has revealed two additional assets: 1) a depicted map of the areas of construction expertise, which is very useful not only to stimulate the communication between people of different institutions working on the same field, but also to identify strengths and weaknesses, and 2) strengthens in promoting communication and information exchange not only making the information more accessible by being independent of the concrete device (laptop, mobile phone, etc), but also making the information more understandable by bridging the language gap between the academic and the industrial community by executing a dynamic mapping process keeping the autonomy of the different input data of these two communities.

Keywords: knowledge base, ontology, OWL, semantic web, inference search.



1 Introduction

The construction sector in the United Kingdom has an output of approximately £102 billion (DTI Construction stats, 2005). Successive governments have commissioned reports on the performance of the sector. All these reports have highlighted the need for radical change in the performance of the sector. The UK construction sector has become increasingly fragmented over the years and more and more work is being subcontracted to specialist, often small, companies. The challenge that still faces the sector is the need to help the very large number of SMEs in construction to improve their performance. The UK Government recognised the importance of this by creating first the 'Movement for Innovation', then 'Rethinking Construction' and the 'Construction Best Practice Programme'. The UK university of Salford commenced the Construction Knowledge Exchange in August 2004. The project has the vision to promote knowledge exchange between industry and universities in the UK construction sector. It attempts to promote and enhance engagement in activities that establish industry needs, capture and share knowledge and build capacity and resource through enabled networks and links that reach out to all levels of business and higher education. The approach presented in this paper lays the foundations for collaborative partnerships between academia and industry in the construction sector, with the aim of developing a strong network of organisations able to share knowledge at a national level and make it available for use at a local level.

2 An ontology-driven approach

To exchange knowledge in the construction sector, i.e. semantic exchange, requires a shared mechanism to classify domain knowledge-items or information into interrelated concepts, i.e. ontology. Ontologies may be used for achieving a common consensus within a user community about conceptualising, structuring and sharing domain knowledge. It is expected that ontologies can help give an integrated interface with the construction sector, which has become increasingly fragmented over the years and has allowed the coexistence of more than one common language. It is recognised that one ontology alone is not going to facilitate the exchange of knowledge between the two identified communities in the construction sector: academia and industry. It is anticipated that an ontology that bridges the two communities will be required, as there are differences between the languages that they use.

This study pursues to obtain an ontology which provides the basic structure for developing a *construction Knowledge Base*. The scope of this *construction Knowledge Base* is to adequately encode the areas of expertise of UK academic community in the construction sector and make this knowledge available for the industrial community. While trying to model an ontology which captures the knowledge gathered from the UK academic community, one soon has to face: 1) 'who has the knowledge', and 2) 'who knows what'. It is not realistic to assume that UK universities which actively work in the construction sector will participate in the integration of the different areas of knowledge to provide a



well-elaborated ontology for their areas of knowledge and expertise. Likewise, it is also not realistic to assume an individual or a small research group of being able to produce and maintain such an ontology. Hence, the approach taken is

- a) To develop a methodology or procedure to acquire the relevant information about the UK academic community and try to formalize the different activities or steps involved as much as possible to facilitate the inclusion of techniques to automate the information acquisition in the future.
- b) To divide the UK academic map into regions, and concentrate in one region each time, where each region has been divided into counties. For example, the North-West region was selected as the initial region to focus on, and it was identified that only the counties of Lancashire, Greater Manchester, and Mersey have Universities which are engaged in activities relevant for the construction sector.
- c) To adopt one of the existing standards available in the construction sector, and slightly update, modify, or extend this standard. This approach requires less effort than constructing the standard from first principles and has the benefit of being already accepted by the existing communities.

2.1 The academic ontology

The current approach pursues to automate as much as possible the creation and maintenance of the *construction Knowledge Base* portion which encodes the expertise of the UK academic community in the construction sector, ideally, the acquired information will be placed in the *construction Knowledge Base* either directly by: a) academics, b) the CKE Administrator (*CKE Admin* for short), or c) automatically by the system. At this stage the information is placed in the *construction Knowledge Base* only by the *CKE Admin*, and the other two alternatives mentioned will be available in the near future.

The current research focuses on the information available online about the UK academic community. A brief description of the three main activities considered to acquire information about the UK academic community is the following:

- 1) *Filtering*: creation of a list of Web pages (URLs), where each URL selected is attached to a UK university/research group/etc which is deeply engaged in activities relevant to the construction sector. This activity has been performed considering different levels of granularity: starting by a macro-level (UK universities) which was refined later at a micro-level (UK research centre/research group/etc) as an insight into each of the UK universities previously selected. In the near future, another refinement will be performed to reach to an academic person-level (Web pages of academic people). The URLs selected are stored in the *Web pages database* and in the future URLs will be added/modify/etc to the *Web pages database* via the *CKE Admin*, who will search for these URLs online and encourage UK universities/research centres/etc to email relevant URLs.
- 2) *Extracting*: processing the content of the URLs selected in the previous activity and extracting the information relevant to the academic *construction Knowledge Base* portion. It is a key challenge of the project to get this



working effectively. The difficulty lies in the wide variety of formats used in home pages and the need to identify the relevant information with an extremely high degree of accuracy from all of the differences between the languages that they use. The extracted information for each URL are stored in the *Web pages database*.

- 3) *Mediating*: fuelling the *construction Knowledge Base* with the information acquired from the URLs selected. The *construction Knowledge Base* stores the academic ontology and its instances.

In construction, domain knowledge has to be supplied on a larger scale. Instead of developing a sophisticated *construction Knowledge Base* from first principles, a 'conservative' approach is proposed – reusing large-scale resources, but refining the data from these resources so that advanced representational requirements imposed by more expressive knowledge representation languages are met. The approach reuses knowledge by integrating concepts and relationships from a generic ontology and a specific ontology obtained by encoding and classifying pieces of knowledge from different URLs.

A core purpose for the use of ontologies is the exchange of data not only at a common syntactic, but also at a shared semantic level. Existing methodologies and practical ontology development experiences have in common that they start from the identification of the purpose of the ontology and the need for domain knowledge acquisition [5], although they differ in their focus and steps to be taken. In this study, to obtain an ontology which provides the basic structure for developing a *construction Knowledge Base*, the three basic stages of the knowledge engineering methodology of CommonKADS [10] coupled with a modularised ontology design have been followed:

I) *Knowledge identification*: in this stage, several activities have been included: 1) exploration of all domain information sources in order to elaborate the most complete characterisation of the application domain, and 2) listing of potential components for reusing. The knowledge sources identified: URLs with factual construction knowledge, and the SWRC ontology [13] which generically models key entities relevant for typical research communities and their relationships.

II) *Knowledge specification*: in this second stage the construction of a specification of the domain model and the underlying ontology has been made. The reusable knowledge selected in the previous stage provided part of the complete specification. Firstly, semiformal modelling of the ontology using the Unified Modeling Language UML [9] in order to represent graphically the objects, class and relations of the domain knowledge. The next activities were carried out: 1) refinement of the reused concepts and relationships, 2) addition of new relationships among the modelled concepts. Secondly, represent the ontology using the Web Ontology Language OWL [15]. Protégé [8] version 3.1.1 has been chosen as the tool, which is an ontology-design and knowledge acquisition tool, to build the ontology in OWL using the Protégé-OWL Plugin. The next activities carried out were: 1) extraction of knowledge from the SWRC ontology [13] in OWL – only some parts of the SWRC ontology were needed – and adaptation of the extracted ontology for the needs of the project, and 2) encode and classify factual knowledge extracted from different URLs selected as

a hierarchy which is integrated with the OWL ontology developed by means of the *ResearchTopic* concept.

III) *Knowledge refinement*: in this third stage, 1) the resulting domain model was validated by paper-based simulation, and 2) contents of the domain model were filled by adding more instances in OWL.

2.2 The industrial ontology

In nearly all ontology-based integration approaches ontologies are used for the explicit description of the information-source semantics, but there are different ways of how to employ them. Constructing a shared domain ontology for the construction industry from first principles is a difficult task. However, several standardisation initiatives have been driven by large consortiums in the construction sector. Over the last decade, projects such as the FUNSIEC project [7] have been conducted or are presently ongoing, establishing classification standards and exchange protocols for the construction industry. The following classification standards and exchange protocols for the construction industry have been reviewed: British Standard 6100 [3], Unified Classification for the Construction Industry [14], Construction Industry Trading Electronically [4], Electronic Business using eXtensible Markup Language [6], Architecture, Engineering and Construction eXtensible Markup Language [1], and Building and Construction XML [2]. From the above list, Uniclass [14] – Unified Classification for the Construction Industry – was chosen as a classification standard for the construction industry. Uniclass is joined by the Construction Confederation, the Royal Institute of British Architects, the Royal Institution of Chartered Surveyors, the Chartered Institution of Building Services Engineers, and the Department of the Environment Construction Sponsorship Directorate. Uniclass as a classification scheme for the construction industry (architecture and engineering) is defined in 15 tables codes for a multi-level international classification of building and civil engineering elements, spaces, documents, phases, materials etc. This construction standard has been selected because: a) it is specific to the construction sector, b) it attempts to be a consensual model of meaning for construction industry, c) it sets out to capture the complexity and fragmentation of construction industry, and d) it is intended to supersede CI/SfB, the most commonly used classification system for construction information which was last revised in 1976.

It was recognised that the Uniclass standard [14] could not be used as it is, and updates, modifications, or extensions of the standard were needed. However, Uniclass is the core of the industrial ontology developed, which has been modelled straightforward as a hierarchy of concepts. The industrial ontology is integrated with the generic part of the academic ontology via the *Topic* concept (one of the reused top-level concepts of the generic SWRC ontology [13]). Hence, the industrial ontology appears as an industrial construction topic hierarchy: *Topic* versus *IndustrialConstructionTopic*. The industrial ontology has been modelled in OWL [15] by means of Protégé [8] version 3.1.1.



2.3 The mapping ontology

The current research proposes a multiple-ontology approach to achieve the Construction Knowledge Exchange. The academic ontology and the industrial ontology have in common the *ResearchTopic* concept (a subtype of the *Topic* concept of the generic SWRC ontology [13]). The specific part of the academic ontology, whose concepts were obtained from factual information extracted from different online sources, keeps its independence from the industrial ontology based on UniClass, and a major challenge towards the Construction Knowledge Exchange is how to bridge the gap between these two specific ontologies.

In [11] has been pointed out that in multiple-ontology approaches the mapping is very difficult to define, because of the many semantic heterogeneity problems which may occur. To overcome the mapping difficulties mapping outlined, the approach relies in an independent construction expert who is familiar with typical tasks and problems in the construction sector, but who is not concerned with a specific information source. The proposed strategy is based on two steps which are inspired in the propose-and-revise method, and that are executed in an iterative process. The individual steps to perform the mapping process between the academic ontology and the industrial ontology are briefly described below:

- 1) *Finding common concepts.* The independent construction expert goes from the academic ontology to the industrial ontology and vice versa to make a semantic matching from one source into another. As a result of a new matching performed a new concept is defined which could be call a “bridge concept”. The most general bridge concepts are “top-bridge concepts” which are the ones that define the mapping performed between the top-level concepts of the academic and industrial ontologies. The “bridge concept” is defined as concretely as possible to obtain the most accurately mapping between the two ontologies.
- 2) *Refine definitions.* The introduced strategy follows an evolving life cycle allowing the independent construction expert to step back all the time to modify, add, and remove ontology mapping definitions, e.g. refining a “bridge concept”.

The mapping ontology has been modelled in OWL [15] by means of Protégé [8] version 3.1.1.

3 Applying the approach

The recent explosion of interest in the World Wide Web has fuelled interest in ontologies. Ontologies are the backbone technology for the Semantic Web and – more generally – for the management of formalised knowledge in the context of distributed systems. In 2004, the Web Ontology Language OWL [15] was adopted by the W3C as the standard for representing ontologies on the Web. As empirically proved by [12] Semantic Web enhanced search accomplishes significantly better search results than other information retrieval techniques. In this sense portals based on Semantic Web technologies, Semantic Web portals



(SW portals for short), represent the next generation of Web portals. For all the reasons outlined, the current approach has developed a SW portal (β -CKE portal) which has ontologies as central components. The β -CKE portal provides three alternatives to explore the content of the *construction Knowledge Base*:

- 1) *UK maps*: explore the different regions and counties to find out the UK universities engage in activities relevant for the construction sector.
- 2) *UK Universities*: explore a classification of the UK universities based on their location, where UK regions are the top level concepts of the hierarchy, and the lowest level concepts are the UK research centres/research groups/etc. In the near future, another level will be added and connected with the actual lowest level: academic person-level. When a UK university/research group/etc is selected, the attributes and relationships of the concept selected will be shown.
- 3) *Construction topics*: explore a classification of the areas of expertise of the academic community whose concepts were obtained from factual information extracted from different online sources and organised as a hierarchy. When a construction topic is selected, the attributes and relationships of the concept selected will be shown.

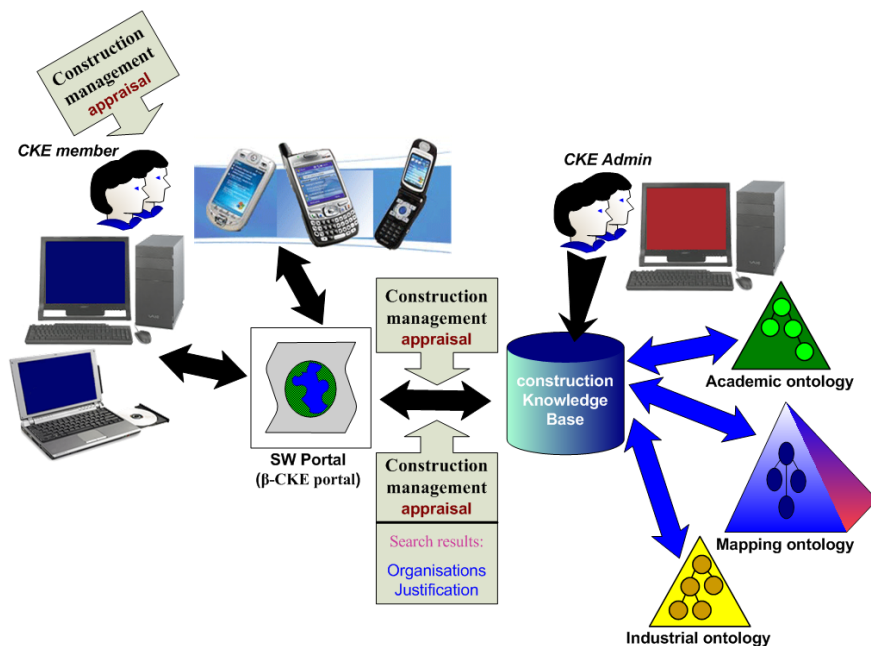


Figure 1: The β -CKE portal.

Figure 1 is an overview of the classical 3 tier architecture that the β -CKE portal incorporates. The figure outlines the following features:

- 1) The classification of β -CKE portal users. Currently only two types of users are considered: a) *CKE Admin* who is allow to access, create, and edit any

information in the β -CKE portal without any restriction, and b) *CKE member* who is a registered user that is provided with a login and a password and can access the information in the β -CKE portal. In the near future, the *CKE member* users will be provided with the appropriate permissions to contribute to the β -CKE portal information provision and maintenance, and another type of user (*CKE guest user*) will be included.

- 2) The devices supported by the β -CKE portal. Two types of devices have been considered to access the β -CKE portal: classical computer-based devices (e.g. laptops), and mobile devices (e.g. mobile phones).
- 3) The search capabilities of the β -CKE portal. An example of use of the inference-powered search engine that the β -CKE portal incorporates is given.

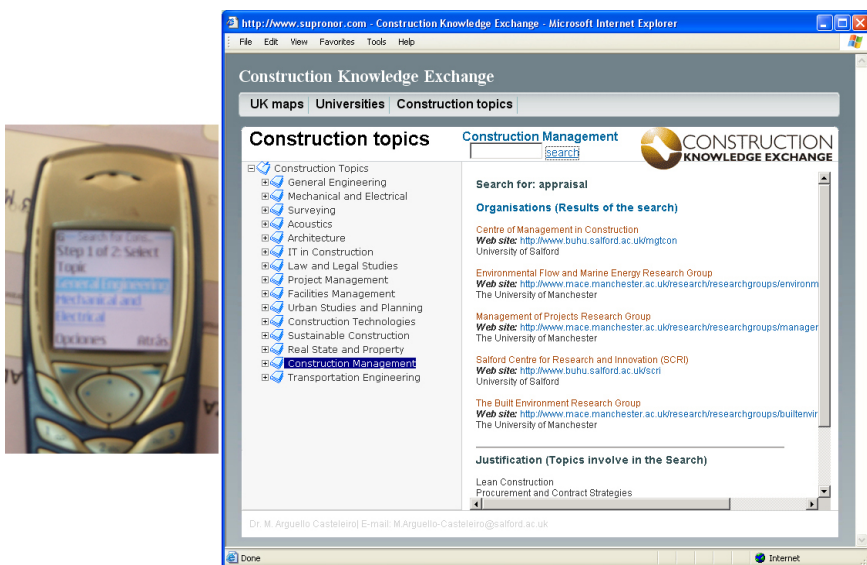


Figure 2: Use of the inference-powered search engine.

Figure 2 shows the two steps that need to be accomplished to use the inference-powered search engine that the β -CKE portal incorporates, in order to demonstrate how the β -CKE portal is exploiting ontologies to a greatest extent:

- 1) Select a top *construction topic* concept from the academic *construction topic* hierarchy (see figure 2 - left).
- 2) Write a word related to the top *construction topic* concept previously selected and press the “search” button. Figure 2 (right) shows the results of the search for the word “appraisal”, where “Construction Management” is the top *construction topic* concept selected.

The inference-powered search engine of the β -CKE portal are intended to be used by people from construction industry, and its main aim is to provide a list of UK research centres/research group/etc with their contact details for any term

about construction that could be addressed to the β -CKE portal; a justification is also provided in order to increase the user's confidence about the answer provided by the system, as all the complexity of the inference-powered search engine is hidden to the user to simplify its use.

4 Evaluation criteria and results

Three of the testable hypotheses list regarding Semantic Web-based information seeking included in [12] have driven the development of a questionnaire for assessing usability of the β -CKE portal.

Table 1: Questionnaire for assessing usability.

| | Question | Answer |
|--------------------|--|--|
| Time on task | 1) Estimate how easy is to use the β -CKE portal to find a University/Research Centre/etc? | A Very easy B Moderately easy C Somewhat difficult D Very difficult |
| | 2) Estimate how easy is to use the β -CKE portal to find the Universities/Research Centres/etc related to a Construction category/subcategory? | A Very easy B Moderately easy C Somewhat difficult D Very difficult |
| Accuracy | 3) How clear is the availability of different options to explore in the β -CKE portal? | A Everything was clear B Mostly things were clear C Some things were clear D Most things were not clear |
| | 4) Estimate how easy was to learn how to explore the different options in the β -CKE portal? | A Very easy B Moderately easy C Somewhat difficult D Very difficult |
| Recall | 5) How helpful is the β -CKE portal to find the Universities/Research Centres/etc related to a Construction category/subcategory or to a construction term/topic? | A Very helpful B Moderately helpful C Somewhat difficult D Not helpful |
| | 6) How clear was the outcome data presented in the β -CKE portal when searching for a construction term/topic? | A Everything was clear B Mostly things were clear C Some things were clear D Most things were not clear |
| Emotional Response | 7) Estimate your level of satisfaction with the β -CKE portal? | A very satisfy B Satisfy C Unsatisfied D Very unsatisfied |
| | 8) When searching for a construction term/topic, the justification of the outcome data provided by β -CKE portal was useful to increase your level of trust to the search results? | A Very useful B Moderately useful C Somewhat useful D Not useful |
| | 9) How comfortable would you be using the β -CKE portal? | A Very comfortable B Comfortable C Uncomfortable D Very uncomfortable |
| | 10) Would you recommend the β -CKE portal to colleagues in the construction sector? | A Definitely recommended B Possibly recommended C Probably not recommended D Definitely not recommended |
| | 11) Comments or Recommendations | |

Table 1 summarizes the questionnaire for assessing usability of the β -CKE portal which has been used for the evaluation sessions performed. The questionnaire has been divided into four areas: 1) *time on task* which is related to the hypothesis number one in [12], 2) *accuracy* which is related to the hypothesis

number two in [12], 3) *recall* which is related to the hypothesis number five in [12], and 4) *emotional response* which is related to how does the person feel about the task completed? i.e. confident?, stressed?.

The *acceptability* of the β -CKE portal is assessed using ten close-ended questions and one open-ended question (see table 1). Close-ended questions elicit feedback in a way that allows easy comparison between the answers obtained from the set of trial interviews performed. The open-ended question is intended to capture the comments or recommendations that could help to improve future versions. Three types of users with diverse backgrounds have been considered to bring complementary knowledge to the project: 1) academic people, 2) industrial people, and 3) computer science people.

So far the evaluation comprises a small number of test users (2 from academia, 4 from industry, and 2 from computer science). The results available are far from having statistical significance, but they are able to illustrate some first impressions about the β -CKE portal. A brief summary of the results obtained is the following: 75% of close-ended questions were answered with letter 'B', 17% of close-ended questions were answered with letter 'A', 8% of close-ended questions were answered with letter 'C', and most of the comments or recommendations provided are intended to suggest alternatives to make improvements that will benefit the use of the system (e.g. increasing the use of alphabetical order). A fairly obvious conclusion from these statistics is that the β -CKE portal is a very promising one. Currently the β -CKE portal is available online, and an increment of the *CKE member* users is an ongoing task to promote massive testing and obtain a more detailed feedback.

5 Conclusions and further work

The use of ontologies is a promising approach in order to capture information semantics and specify a shared vocabulary, however, a strategy is needed to determinate what kinds of ontologies are needed and how they can be built and integrated. Beside the mapping problem, a striking lack of sophisticated methodologies supporting the development and use of ontologies has been found. The development of a strategy was a major step in the ontology-driven approach presented in this paper, to bridge the language gap between the academic community and the industrial community in the construction sector, which has become increasingly fragmented over the years and has allowed the coexistence of more than one common language. The illusion of an integrated view from the user's perspective is achieved by executing a dynamic mapping process keeping the autonomy of the different input data of these two communities. The inclusion of an academic person-level as well as the availability of an electronic bulletin board in the near future will foster the creation of collaborative partnerships between academia and industry in the construction sector. The *acceptability* of the β -CKE portal, which has ontologies as central components, is assessed by the answers obtained from the set of trial interviews performed using a questionnaire for assessing usability. So far the evaluation comprises a small number of test users, although the current



availability of the β -CKE portal online will promote massive testing and obtain a more detailed feedback.

Acknowledgements

We thank to the following Spanish organizations: *ACS Sistemas*, and *Consortio Zona Franca Vigo* for their friendly collaboration and fruitful discussion. In particular, we thank *Angel Avalos Boado*, *Florindo Alvarez Perez*, and *Ana Rosa Villarino Lopez* who contributed essential parts of this work presented.

References

- [1] aecXML, <http://xml.coverpages.org/aecXML.html>.
- [2] bcXML, <http://xml.coverpages.org/econstructSummary.html>.
- [3] BS6100, <http://www.idef.com/idef0.html>.
- [4] CITE, <http://www.cite.org.uk>.
- [5] Davies, J., Fensel, D. and van Harmelen, F. editors. (2002) *Towards the Semantic Web: Ontology-Driven Knowledge Management*. John Wiley.
- [6] ebXML, <http://www.ebxml.org/>.
- [7] FUNSIEC project, <http://www.funsiec.org>.
- [8] Protégé, <http://protege.stanford.edu/>.
- [9] Rumbaugh, J., Jacobson, I. and Booch, G. (1999) *The Unified Modelling Language Reference Manual*, Addison-Wesley.
- [10] Schreiber, A., Akkermans, H., Anjewierden, A.A., Hoog, R., Shadbolt, N.R., Van de Velde, W. and Wielinga, B. (1999) *Engineering and managing knowledge. The CommonKADS methodology*. The MIT Press.
- [11] Stuckenschmidt, H. and van Harmelen, F. (2005) *Information Sharing on the Semantic Web*, Springer-Verlag, Berlin, pp33.
- [12] Sure, Y. and Iosif, V. (2002) *First results of a Semantic Web Technologies Evaluation*, Proceedings of the Common Industry Program, University of California, Irvine, October.
- [13] SWRC ontology, <http://ontoware.org/projects/swrc/>.
- [14] Uniclass, <http://www.productioninformation.org/Uniclass.html>.
- [15] W3C (2004), OWL, <http://www.w3.org/2004/OWL/>.

