Ontologies in the CRP Henri Tudor projects

Jean-Sébastien Brunner and Thibaud Latour

Centre de Recherche Public Henri Tudor - Centre for IT Innovation (CITI),
29 Av. John F. Kennedy – L-1855 Luxembourg,
{jean-sebastien.brunner,thibaud.latour}@tudor.lu,
http://www.citi.tudor.lu

Abstract. The Centre de Recherche Public Henri Tudor is as a public research centre whose main objective is to promote innovation in both private and public sectors. The purpose of this poster is to present briefly the expertise of the CITI, the department of information and communication of the CRP Henri Tudor, in the domain of ontologies through an overview of some relevant projects.

1 Ontologies for Knowledge Management

1.1 Dynamic Knowledge Repository to Build State-of-the-Art

In the *Capital* project, the CRP Henri Tudor developed a capitalization tool for formalization projects[1]. The main objective of formalization projects, in the context of innovation and technology transfer, is to gather a critical set of state-of-the-art knowledge pertaining to a given engineering domain, and to provide a structure which makes the formal description of this knowledge transferable so that the recipient is able to acquire a new expertise in the domain.

The aims of this tool consist in capitalizing the knowledge acquired in research projects and keeping up to date the documentation, which can be automatically published. The main characteristic of this tool is to let the user collectively and progressively build a knowledge base designed to formulate a particular engineering domain. In order to construct this knowledge base, one has to build the ontological structure of domain knowledge by instantiating a core meta-ontology. The core meta-ontology comprises five categories of object classes:

- **Static objects**, not subject to significant change during time and independent from their context of application (Tool, Method . . .);
- **Transformation objects**, describing transformations of things from one initial state to a final state(Process, Transformation activity);
- **Reference objects**. They are bibliography reference, glossary, URL . . . ;
- **Contextual objects**, which depend on time and context of use (Case study, Problem resolution, . . .) ;
- **Semantic objects**, used to express a richer semantics.

Once the ontology is structured, an interview mechanism is generated to help users populating the knowledge base dynamically, according to the current state of the ontology. This process allows users to interact with the system without being specialist in knowledge engineering. The ontology is flexible, allowing regular updates of the structure through a totally similar interactive scheme as the one used to populate the knowledge base.

1.2 Semantic Annotation in Scientific and Technological Watch

Ontologies can be used as a core element of an electronic document management system. In the context of Scientific and Technological Watch (STW) information management, ontologies are used to model a multidimensional thematic space [2]. Document content is modelled by a combination of several concept vectors, representing each a subset of concepts belonging to a particular point of view. The thematic space is defined by the Cartesian product of these concept vectors.

Beyond this thematic space itself, each dimension can adopt different structures, depending on the required degree of expressivity. The highest degree of expressivity is obtained in structuring each dimension by an ontology. Searches can hence benefit from all relationships linking these concepts in the ontology structure. The thematic map model makes possible the computation of a similarity measure (used to retrieve documents according to user interests) based on fuzzy set theory. In order to avoid spurious propagation of mapping function between documents and the reference ontology through hypernymy relations, we introduced fuzzy membership functions in the ontology meta-model.

Finally, this model can be used for content modelling of other kinds of documents and not only in the context of STW. For that, dimension-set need to be adapted to the new context. As a consequence, new meta-ontologies must be constructed with the help of a tutorial we are currently writing.

2 Ontologies for Domain Structuration

2.1 Business Model in Efficient Project

Efficient (E-business Framework For an effIcient Capture and Implementation of ENd-to-end Transactions) aims to identify e-business transactions offering an important value-added, and to make a logical representation of transaction content (including message and rules) in order to check up the transaction feasibility prior to its implementation [3].

For that, the Efficient project defines the business model as an ontology. This one is used to store the structure of the model, and the ontology flexibility enables evolutions of the structure when model is upgraded. This model is derived from *business model ontology for formalising e-business strategies* (developed by Osterwalder and Pigneur of HEC Lausanne), completed with references in order to be used in real enterprise cases. Beyond modelling, this ontology permits the analysis of business models and suggests diagnostics thanks to the data interpretation it enables.

2.2 RIO : A Structure for Research Information in Luxembourg

In Luxembourg, Research and Innovation resources are federated in a governmental portal (www.innovation.public.lu). In collaboration with Luxinnovation, the CRP Henri Tudor is in charge of exploring potentialities for the semantic extension of the portal. For that, we built an ontology for representing the main activities in the research area in the Grand Duchy of Luxembourg.

The Research and Innovation Ontologies (RIO) bears two main purposes: the first one is to support all data exchanged between innovation actors in the country, and the second one is to provide an advanced and processable knowledge structure (by specifying relationships between classes) to add semantic capabilities to the R&D portal. These capabilities mainly consist in advanced and graphical browsing within data and also intuitive advanced searches.

2.3 TAO, An Adaptative Testing Platform

TAO is a flexible and distributed platform for administrating, authoring and delivering personalised multimedia tests [4, 5]. In this project, all data are described by a variety of ontologies: subjects of tests, groups of subjects, tests, items, etc.

Entirely created by the users and stored in RDF, knowledge structure is not known a priori. As a consequence, users can define not only data but also data structure. Moreover the whole application is totally flexible, letting structure dynamically evolve. TAO is built as a distributed architecture in which each module has its own structure. The interoperability between different modules or users is provided by RDF triplet exchanges when users request either the structure or the content of a remote ontology if their credentials allow it.Two additional aspects are managed by the ontology: the graphical user interface, and the access rights.

Finally, tests and item structures were described in XML/RDF and transformed in Macromedia Flash for execution in the client browser. Further development aims at describing not only the data structure in RDF but also the behaviours and algorithmics of testing. These capabilities are expected to be provided as a result of the aforementioned projects.

3 Research Areas around Ontologies and Semantic Web

3.1 Collective Knowledge Structuration

As part of the project Semantic R&D, we are setting up a method to collectively build ontologies. Our philosophy aims at minimal interference with daily work of users while trying to make the structure of the knowledge they use emerge from their activities. To do so, we designed a bottom-up method which is able to federate personal data organization into an organization ontology. Our method relies on the use of personal logical desktop organizers which structures are sharable among a community.

After a period of use, it is expected that substantial exchanges have been performed that yields a partial common knowledge structure from individual views. The next operation consists in formalizing the common structure. This requires knowledge expertise and advanced techniques of ontology mapping. This process is experimented in different scientific communities.

3.2 The Semantic Web: a Basic Open-World Infrastructure for Processing Knowledge

In 2004, the Semantic Web (SW) is still a research area with only few applications. To answer the question "why" we assume that there are no tools to browse the SW, like Internet Explorer or Mozilla browse the classical WWW. The main difficulty is that SW should be both human and computer readable. As such, the SW can be viewed as a huge knowledge base spread over Internet servers. Contrary to distributed databases, there is no global integrity and data represent knowledge and should thus contain all necessary semantics to be processable on its owns.

From this, we devised a generic tool able to read and interpret the SW: a semantic web processor. This processor should be able to load any data, the description of operation you can make on it, and will give response to a request on these data. One of the main difficulties lies in the fact that the processor is very generic and should ideally be able to work with the least a priori-knowledge on data meaning and structure. As a consequence, language for data description will not be defined in the processor core code but should be brought as another data. We are currently exploring the description of behaviour (execution capabilities) and implementing a prototype of this processor in SWI-Prolog.

References

1. Latour, T., Incoul, C., Brimont, P., Leidner, S.: Framework for Case Study Engineering (Fracasse). Documentation Package, v1.0. Technical report, CRP Henri Tudor (2002)
2. Brunner, J.S., Latour, T.: Referencing Text Documents in Multidimensional Concept Spaces for Technological and Scientific Watch. In: Proceedings of the Workshop on Terminology, Ontology and Knowledge Representation, 22-23.01.2004, Lyon, France. (2004) 11–18
3. Eshuis, R., Brimont, P., Dubois, E., Grégoire, B., Ramel, S.: EFFICIENT: A Tool Set for Supporting the Modelling and Validation of ebXML Transactions. In: Proceeding of the joint 9th European Software Engineering Conference (ESEC) & 11th SIGSOFT Symposium on the Foundations of Software Engineering (FSE-11), ACM Press (2003) 359–362
4. Busana, G.: Comment articuler l'efficacité et l'équité d'un système éducatif plurilingue ? In: 16e Colloque International de l'Admée-Europe Evaluation: entre efficacité et équité, 04-06.09.2003, Liège, Belgium. (2003)
5. Martin, R., Busana, G., Latour, T., Vandenabeele, L.: A distributed Architecture for Internet-Based Computer-Assisted Testing. In: Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications ED-MEDIA 2004, 21-26.06.2004, Lugano, Switzerland. (2004)