

Semantic UBL-like documents for innovation

R. Volkan Basar¹, A. Anil Sinaci¹, Fabrizio Smith², Francesco Taglino²

¹SRDC Software Research & Development and Consultancy Ltd.
Silikon Blok No:14, Teknokent ODTU, 06800 Ankara, Turkey
{anil,volkan}@srdc.com.tr

²National Research Council, IASI “Antonio Ruberti”
Viale Manzoni 30, 00185 Roma, Italy
{smith,taglino}@iasi.cnr.it

Abstract. Considering innovation as the result of only spontaneous activities is a simplistic vision, because working out inspiration and reaching up to innovation need awareness and knowledge about the application domain and its problems. In this paper, we address the issue of knowledge representation, access and sharing in an enterprise context, by proposing an ontology-based framework (DocOnto) for the semantic description of documents involved in innovation activities. The framework, which is built within the BIVEE European project, is characterized by a customizable approach inspired by the UBL/CCTS, which allows each enterprise to refine the DocOnto at best for its needs. Then UBL-like structures are semantically lifted and used for describing concrete documents. Such a semantic representation enables reasoning services like querying and retrieving of documents, understanding similarities among documents, assessing their status and quality, monitoring innovation activities. The framework is supported by the technological integration of the iSurf eDoCreator, for modelling UBL-like documents structures, and the Production and Innovation Knowledge Repository (PIKR), the semantic knowledge hub of the BIVEE platform.

Keywords: business innovation, ontologies, semantic description, UBL, CCTS, document management.

1 Introduction

“Genius is one percent inspiration and ninety-nine percent perspiration”. This quotation from Thomas. A. Edison reveals what is behind innovation successes. Innovation is often identified as the result of spontaneous attitudes like creativity or artistic flair, but bringing new ideas into the market in the form of innovative products or services, also needs the adoption of methods and supporting means. In particular, in the era of the information society, knowledge management has a primary role also in relation to innovation activities. With respect to that, [1,4] consider innovation as a practice and process that captures, acquires, manages and diffuses knowledge with the aim to create new knowledge. Furthermore, knowledge enables creativity by

permitting knowledge associations and linkages that otherwise are difficult to be discovered.

In this paper, we outline a framework for designing semantics-based structures (Document Ontology, DocOnto) to enable the semantic enrichment and management of innovation-related documents, i.e. documental resources produced and *consumed* during innovation initiatives (e.g., proposed ideas, feasibility studies, etc.). This activity is conducted within the BIVEE¹ European project, which is about the development of an ICT infrastructure for supporting innovation activities in virtual enterprise (VE) environments.

Among related initiatives, we mention Dublin Core², a vocabulary of fifteen properties for description of documental resources, and SALT [8], which is for describing the organization of a document in terms of sections and paragraphs. While we intend to re-use part of the terms from Dublin Core, we look at documents differently from SALT, since we are focused on the semantics instead of the organization of the structure of a document.

The proposed framework is based on the one hand, on the methodological integration of the UBL/CCTS [7] approach, which is for modeling and customizing documents structures, together with semantic representation methods. On the other hand, the framework is supported by the technological integration between the iSurf eDoCreator[6] UBL editor and the Production and Innovation Knowledge Repository (PIKR), which is the semantic knowledge hub of the BIVEE platform. The objective is the semantic lifting of innovation-related documents structures and content for enabling interoperability and openness, as well as reasoning services such as querying and retrieval of documents, reasoning over documents description, understanding similarities among documents, assessing status and quality of documents and, monitoring innovation activities.

The paper is organized as follows. Section 2 presents the overall structure of the DocOnto and the UBL-inspired approach for building and customizing the DocOnto itself. Section 3 focuses on the technical aspects concerning the integration between the PIKR and eDoCreator, as well as on the services provided by the PIKR for reasoning over the semantic description of annotated documents. Conclusions and future work end the paper.

2 The DocOnto framework

One of the main objectives of the BIVEE project is to support and facilitate innovation activities in a VE environment. To this end, the Virtual Enterprise Modeling Framework (VEMF) has been developed. According to the VEMF, innovation-related activities happen within four waves: Creativity, Feasibility, Prototyping and Engineering. Flowing through these four waves, many documents are produced, used, consumed and evaluated. For instance, in the Creativity wave, given a problem or issue, many ideas can be proposed to address it. Some of them will pass

¹ Business Innovation and Virtual Enterprise Environment (No. FoF-ICT-2011.7.3-285746).

² <http://dublincore.org/documents/dces/>.

the initial stage and will be further elaborated. Recording such information means keeping track of reasons that guide decisions, and re-using knowledge to save time and money in the future. We consider that ontology-based semantic technique can be effective in addressing representation, sharing, access, and reasoning over documental resources, especially in VE context where boundaries are larger and a reference (ontology) is requested.

For the definition of the proposed ontology-based innovation document framework (DocOnto) we started from the results of an activity performed within the BIVEE project: two end-user organizations were asked to see their innovation-related activities through the four innovation waves and indicate the information they actually produce and use. This brought the identification of sets of documents, one for each end-user [5]. These results have been taken as specifications and, starting from them, a conceptualization of these documents has been performed for identifying valuable *InfoItems* (building blocks, which correspond to small and meaningful elements), *InfoSets* (recursive aggregation and association of *InfoItems*) and associations between them as described below and reported in Table 1.

- **Header** groups meta-data *InfoItems* like the title of the document, the authors etc.
- **Content** groups *InfoItems* describing the essence of the document, i.e., its semantics. The adoption of domain-focused dictionaries, thesauri or ontologies increments the level of interoperability and enables reasoning mechanisms.
- **Related Knowledge Resources** section allows to establish relations between *InfoSets* such as 'prerequisite', 'feedback', 'partOf', 'relatedTo' etc.

Header	
Title	Advanced HMI
Description	System for the robot programming based on the 3d reconstruction of the inspected components
...	...
Content	
Research Line	3D vision, cloud point, artificial intelligence algorithm
Technology	HMI
...	...
Related Knowledge Resources	
Part of	doc:IP_AdvancedHMI

Table 1. An example of *InfoSet* instance: a Technical Solution Report

2.1 UBL Customization Approach

The UN/CEFACT Core Component Technical Specification (CCTS) has the notion of building blocks, called Core Components (CC). Core components are context-neutral having a generic semantics and purpose, and can be re-used in different contexts [6]. Business Information Entities (BIE) are contextualized CCs and have three types: Basic Business Information Entity (BBIE), Association Business Information Entity

(ASBIE) and Aggregate Business Information Entity (ABIE). UBL [7] implements CCTS and publishes XML based Business Document Definitions, Common BIEs and Data Types such as an Invoice document or an Address BIE

Data requirements change for different virtual enterprises in order to address the needs of innovation activities. Hence, it is required to customize the DocOnto for each virtual enterprise once the requirements have been set. UBL provides a methodological way for the customization of already available documents and BIEs. Since this methodology has already been implemented by eDoCreator tool, our solution inherently supports customization of existing innovation related documents and BIEs. According to the UBL standard, new information entities can be added to meet the requirements of a specific business context, optional information entities can be omitted, the meaning of information entities can be refined, new constraints can be specified, new aggregations or documents can be combined or assembled or new business rules can be added during a customization. If a new type of innovation document is required, users can model its structure through customization facilities offered by eDoCreator, which are conformance with customization guidelines of UBL. Since we model documents through *InfoItems* and *InfoSets*, and follow the UBL approach, our modelling directly maps to UBL terms when we leave out the technologies of our framework. This mapping can be depicted as follows: BBIE - *InfoItem*, ABIE - *InfoSet* and ASBIE - *Associations*.

3 TECHNICAL REALIZATION

In this section we give an overview of the technical aspects related to the integration of the PIKR [2], and the eDoCreator for supporting the implementation of the DocOnto. We also outline the semantic services in charge of exploiting the semantic description of the documents in terms of the DocOnto.

About the integration between the eDoCreator and the PIKR, the former exports XML Schema³ of modelled documents to a *Mediator* module. The *Mediator* performs the semantic lifting by encoding documents structures into OWL/RDF⁴, the de-facto standard for ontology and meta-data sharing. The result of the lifting is then transmitted to the PIKR, which maintains it in a triple store.

The knowledge representation framework discussed in the previous sections enables the enactment of a number of reasoning facilities to support the management of documents in innovation projects, in terms of the following services.

Search. This service provides keyword-based search functionalities. The user request is expressed as an ontology-based feature vector describing the criteria for the selection of the resources of interest. By applying semantic similarity techniques (the *SemSim* metrics [3]) the degree of matching among the terms used to formulate the request and the ones used to describe the available resources is computed, and a list of ranked results, with respect to the *Semsim* similarity metrics, is returned. For instance,

³ <http://www.w3.org/TR/xmlschema-0/>

⁴ <http://www.w3.org/TR/owl2-overview/>

suppose that the user is interested in finding all the documents that have been authored in the last two years and concerning the initial stages of the design of a piece of furniture equipped with an electronic device. The corresponding request should be formulated as follows:

```
{content:[Furniture, Electronic_Device]; type=Proposal,
      creationWave=Creativity, issueYear>2010}
```

The engine will retrieve semantically related resources, such as Proposed Idea or Project Proposal documents about a Contour Chair with an embedded Media Player (which are assumed to be defined in the domain ontology as kinds of piece of furniture and electronic device, respectively).

Query. This service enables us to retrieve pieces of knowledge which exhibit some given properties. Queries are posed in terms of the vocabulary and semantic relations provided by the PIKR ontologies, and the underlying reasoning engine returns a list of answers that satisfy all the specified properties. These answers may consist of factual knowledge (DocOnto instances), conceptual knowledge (ontological terms), or references to concrete resources. We are currently developing a query language, based on SELECT-WHERE paradigm along the line of the SPARQL⁵ standard. For instance, to identify reusable best practices or technical solutions in a given domain, we may want to retrieve all the protocols related to documents addressing the research line *3D_Vision*. This can be expressed as follows.

```
Q(?p) : protocol(?p) AND related(?p,?doc) AND
research_line(?doc,3D_Vision)
```

Compliance Checking. This service allows us for checking the compliance of the factual knowledge, captured at a given time in the semantic description of the documents, with respect to business policies and internal regulations. Compliance requirements can be represented in the DocOnto as business rules, i.e., statements that define or constrain the structure of the documents or the dependencies among them on the basis of the sequencing of business operations. The compliance check verifies the consistency among the assertions contained in the DocOnto instances and the axioms defined in the Knowledge Resource Ontologies formalizing the business rules. Examples of constraints are "Each Innovation Report needs to be composed by a Project Proposal and a Market Analysis", or "A Monitoring Sheet cannot be produced unless a Gantt Chart has been finalized before". The former rule can be formalized by the following axiom:

```
if innovation_report(x) then ∃ y,z. project_proposal(y)
and market_analysis(z) and partOf(x,y) and partOf(x,z)
```

⁵ <http://www.w3.org/TR/rdf-sparql-query/>

4 Conclusions and Future Work

In this paper we outlined an ontology-based framework for semantic description of innovation-related documents. We have elaborated on CCTS and UBL approaches and identified a bunch of *InfoSets* corresponding to categories of information that are produced, consumed and evaluated during innovation projects. Furthermore, we identified relationships that can occur among *InfoSets*, and we started to identify *InfoItems*, elementary components of the *InfoSets*. We intend to re-use available vocabularies as much as possible to enable Linked Data approach in this document management methodology.

References

1. Cohen, W.M. and Levinthal, D.A. (1990), "Absorptive capacity: A new perspective on learning and innovation", *Administrative Science Quarterly*, Vol. 35, 128-152.
2. Diamantini, C., Potena, D., Proietti, M., Smith, F., Storti, E., Taglino, F.: A semantic framework for knowledge management in virtual innovation factories. *International Journal of Information System Modeling and Design*. To appear.
3. Formica A., Missikoff M., Pourabbas E., Taglino F. (2013) Semantic search for matching user requests with profiled enterprises. *Computers in Industry*, 64: 191-202.
4. Gloet, M. and Terziovski M. (2004), "Exploring the Relationship between Knowledge Management Practices and Innovation Performances", *Journal of Manufacturing Technology Management*, Vol. 15 No. 5, pp.402-409.
5. Sinaci A., Piersantelli M., Cristalli C., Gigante F., Laleci G., Basar V. (2012), "A Document Centric Approach for User Requirements in BIVEE", *CEUR Workshop Proceedings Vol. 864 Article 5*
6. Tuncer F., Dogac A., Postaci S., Gonul S., Alpay E. (2009), "iSURFeDoCreator: e-Business Document Design and Customization Environment"
7. OASIS UBL TC (2006), "Universal Business Language v2.0". Retrieved March 3, 2013 from <http://docs.oasis-open.org/ubl/os-UBL-2.0/UBL-2.0.pdf>.
8. Groza T., Handschuh S. (2009), "Salt Document Ontology (SDO)". Retrieved March 3, 2013 from <http://salt.semanticauthoring.org/ontologies/sdo#>.