Extracting a semantic view from an ebusiness vocabulary

Nicola Gessa  
University of Bologna  
gessa@cs.unibo.it

Matteo Busanelli  
ENEA, Bologna – Italy  
matteo.busanelli@bologna.enea.it

Piero De Sabbata  
ENEA, Bologna – Italy  
desabba@bologna.enea.it

Fabio Vitali  
University of Bologna  
fabio@cs.unibo.it

Abstract

B2B applications need common formats and standards upon which to implement interoperability tools; in this paper the authors describe an approach to provide a semantic description of a document-based collaborative framework. The starting point is an ISO 11179 compliant vocabulary of business terms and a set of processes and documents, represented using a modular ontology expressed in OWL; the ontology is extracted by a software module in an automatic manner. The authors expect the ontology will support the maintenance of the framework and the data integration toward enterprises. This approach has been applied on a standard dictionary of the Textile/Clothing sector and represents the basis for the development of an ontology, named ONTO-MODA, within the European research project LEAPFROG IP.

1. Introduction

The adoption of ICT solutions for ebusiness can result critical in the scenario of global competition. In this context, the system heterogeneity to compose could discourage the integration process between enterprises involved in the supply chains. The effort to tackle this issue takes basically two approaches: the development of new interoperability software architectures, and the raise of international standardisation initiatives. Standards are often too difficult to develop and to adopt[9]; a mechanism to improve standard usability, maintenance and efficiency could reside in the formal description of the semantic of the standards. In this perspective semantic descriptions and the bounded technologies are intended as a support for standard management, and not as a substitute for them. In this paper we describe our approach to provide a semantic description of a document-based interoperability framework for business collaborations.

2. State of art

The last research developments [5][6] about interoperability have highlighted the relevance of the semantic aspect in this field[4][7]. Semantic descriptions are particularly relevant in the context of the document-based frameworks that underpin upon the definition of a set of components used to build business document templates; a list of the most relevant frameworks includes the CommerceNet eCO framework, ebXML, RosettaNet, UBL. The study of semantic aspects for interoperability joins with the Semantic Web vision, where machines can understand the semantic of the documents diffused in the web without human assistance; in the e-business context, formal descriptions could be provided through the definition of an ontology that represents that implicit concepts and the relationships that underlie the business vocabulary, and consequently the documents. Many research prototypes are now exploiting ontologies and semantic web technologies to improve
interoperability. In [10] the idea is to translate the XML Schemas of the exchanged business documents into OWL descriptions. Then the different ontologies will be integrated using the RACER reasoner in a unified merged ontology that can then be used as a source for the mapping between the different documents.

In [11] interoperability is achieved exploiting an ontology constituted of three layers: the Upper Domain Ontology, the Application Ontology and the Lower Domain Ontology. The idea is to provide, using transformation rules, mapping between the local schemas of a system and the common ontology. Also in [12] is presented a multi-layers architecture that adopts ontologies as a mean to provide semantic mapping in a very similar way as [11], the interoperability between different frameworks is achieved thanks to a transformation process that led to a semantic model of the documents (using OWL) and exploiting a unified document ontology.

Finally, [3] describes a ontology-based approach for a semi-automatic integration of heterogeneous data-sources. The proposed framework exploits the definition of a Global Virtual View (GVV), that represents the common ontological description of the target domain, as the medium to access the data-source.

3. From the business dictionary to the ontology
Our approach building an interoperability framework was designed around two main aspects: the definition of a B2B protocol and of a set of tools that provide the functionalities needed to exchange business information exploiting the defined protocol. To this aim, we have defined a vocabulary of business terms tailored for the target sector (see section 4), together with a software architecture for the management of the framework itself [1]. The vocabulary has been designed following the ebXML model and developing a set of Basic Information Entities (BIE) and Aggregated Basic Information Entities (ABIE) to structure the vocabulary in a hierarchic manner. The vocabulary maintains the definition of the XML elements, their XSD types and the structures of the documents [2].

Collaboration frameworks have moreover to manage also the business logic [7]; considering this critical aspect, we enriched our vocabulary with the definition of the activities and the processes that can be implemented by the business partners. The whole structure of the vocabulary has then been designed to maintain information about three main knowledge domains, each of which is fundamental to give raise to a standardisation initiative:
- Structures of the documents.
- Collaboration scenarios.
- General concepts treated by the framework.

We have now 1) integrated our framework with the semantic description of each of these domain 2) exploiting a software component (developed within our software architecture) that can extract in an automatic manner the needed information from the vocabulary. These domains basically represent three semantic areas to include within an ontology.

In the field of ontology development it does not yet exist a well established design approach: it in general results to be an iterative process, where the ontology is continuously evaluated and improved.

In our approach, we started identifying the basic concepts we want to extract from the vocabulary. Then we created a draft ontology to identify the best structure to model the final ontology. Finally, we have developed a tool that build automatically the whole ontology following the pattern fixed in the draft; for each of the listed semantic areas, we identified specific basic concepts used as cornerstones for the ontology.

We have then mapped these basic concepts to a set of OWL superclasses; they represent the roots of a set of corresponding sub-ontologies that describe these superclasses together with their subclasses and their relationships with other classes or sub-ontologies; each of these basic concepts represent also a starting point for the browsing of the ontology.

For example, we mapped the basic concept of Process in the superclass BusinessProcess and in the corresponding sub-ontology BusinessProcess (that is described in the BusinessProcess.owl file).

With this approach, we have generated a modular ontology, in which each basic concept (and the associated sub-ontology) can be managed independently from the others and is identified by its own namespace; we have moreover streamlined the designing process and the structure of the ontology, easing its maintenance and its future development.

The following sections deepen the content and the structure of the semantic areas listed above.

3.1 The process area
This area models the business scenario of the target business sector. It includes the following concepts:
- Process: A generic process of the supply chain
- Activity: A generic activity within a process
- Document: A generic business documents exchanged between two actors.
- Actor: A partner involved in a business process
- Good: The good treated in the business process
Each of these concepts is then specialised in subclasses. Each process, activity or document is characterised by a type, and refers to production good; the ontology represents the taxonomy of these concepts, together with the relationships between the processes, the activities and the documents. Finally, each document is linked with a component that represents the root of the XML documents.

In this way we connect the world of the business scenarios with the world of the business document.

### 3.2 The document area

This semantic area defines the structure of the vocabulary terms used within the business documents (in the following we call them components), and consequently reflects also the structure of the business documents themselves. The designed ontology does not model the syntax of each document: this task is achieved through proper XML Schemas (see [6]); we have rather decided to highlight the relationships between defined XML types and the semantic content of the documents (to this aim this ontology exploits the component ontology described in 3.3).

This means also that those layout elements of the documents, (like “Header” or “Body”) used for presentation or formatting purposes are not included in the ontology. The basic concepts of this area are:

- **Document**: A generic business document exchanged between two actors.
- **Component**: An XML element or attribute used by the documents.
- **Type**: A generic XML Schema type.

We basically model the idea that a component, regardless it is an element or an attribute, is an instance of an XML type. We have then created the class Component to represent a generic component (element or attribute) that can be used to build a document. Each component maintains a relationship with the class XMLType. We thus maintain separately both the XML types and their instances. Finally, each document is linked with a specific component that represents the XML-root of the document.

### 3.3 The vocabulary area

This semantic area contains the generic concepts and properties that represent the “world” described by the vocabulary and bounded with the XML components. An ISO/IEC 11179[8] compliant vocabulary basically maintains a set of Data Element, each of them composed of its Object Class, Property Terms and Representation Terms; the three dimensions used to characterise our components are then:

- **Object**: An abstract concept related to the business.
- **Property**: A property of an object, described by a component.
- **Component**: the component (an XML element or attribute) used by the documents; the Representation Terms is given by the types of the components.

For example, our vocabulary contains the component (that finally is mapped to an XML element) “FabricCompos”. This component is used to describe the property “fibrous composition” of a “fabric”. In the vocabulary, the component is then directly related with the object “Fabric” and with the property “fibrous composition”. In this way, the framework maintains a list of triples Object-Property-Component.

In fig. 1 is depicted the final structure of the ontology: it shows the three semantic areas and the subdivision of the areas in sub-ontologies (identified by a name); the links between the sub-ontologies (i.e. the defined relationships) are represented by the arrows. The sub-ontologies are connected defining for the superclasses some ObjectProperties that have their domain local to their sub-ontology, and their range defined in an external sub-ontology.

The image highlights some overlaps between the semantic areas: for example, the BusinessDocument class appears both in the Process Area and in the Document Area. These overlaps represent the contact points of the different semantic domains of the ontology, allowing to interconnect all the sub-ontology in a common vision of the framework.

### 3.5 The implemented architecture

Once we have defined the structure of the target

![Figure 1. The global ontology structure](image-url)
ontology, we developed a software tool to generate in an automatic manner the OWL files starting from the information held in the vocabulary. This tool is used each time a new version of the framework is realised, and thus works as a batch application. It has been developed upon the Java architecture J2EE; in particular, we have exploited the Protégé OWL-API developed by Holger Knublauch at the Stanford Medical Informatics(SMI), and adopted by the OWL Plugin of Protégé.

Our ontology is not thought to work in isolation respect other knowledge domains. In this way we have identified two initiatives for the definition of middle and upper level ontologies that can be exploited to integrate different semantic descriptions:

- The SUMO ontology, that defines very general concepts, like time and number.
- The MILO ontology, that defines more specialised concepts (like product, order or delivery).

Our set of the basic concepts (process, good or documents), mapped in the superclasses, and the concepts enclosed in the vocabulary area, represent our interface towards middle level ontologies like MILO. OWL provides some constructs that can be used to associate our concepts with those defined in an external ontology: for example equivalentClass or equivalentProperty, seeAlso or sameAs.

4. Conclusion and future work

The work presented in this paper has been undertaken within an effort for the definition of an interoperability framework for the Textile/Clothing sector, that feels really critical the interoperability issues, also for the large presence of SMEs. This effort originated within the MODA-ML project (www.moda-ml.org). Its results were included in the European standardisation initiative TexSpin, promoted by Euratex and by CEN/ISSS. The MODA-ML group is now improving the framework, supporting the industry adoption through the participation to many activities and to the CEN/ISSS TexWeave initiative. The last version of the vocabulary is composed of 494 terms, 32 business documents, and specifies 8 business processes. The generated ontology includes 373 classes, 44 relationships and 1098 instances.

The main outcomes of our efforts is the definition of an architecture for the management of the semantic of an ISO11179 compliant business vocabulary. Since ISO11179 is the basis of many standardisation initiatives (like UBL) this result opens the way to create domain ontologies from formalised and standardised vocabularies. Our results represent also the core around which we are now developing an ontology, named ONTO-MODA, within the European research project LEAPFROG IP. These results represent also a bright premise for further developments in three areas:

- mapping and integration mechanisms to interface each other heterogeneous data models;
- study the definition of business vocabularies starting from an ontology;
- linking of the ontology with other external ontologies.

7. References