Providing a semantic description for an interoperability framework using ontologies.

Piero DE SABBATA¹, Nicola GESSA², Matteo BUŞANELLI¹, Arianna BRUTTI¹, Angelo FRASCELLA¹
¹ENEA, via Don Fiammelli 2, Bologna, 40141, Italy,
Tel. +39 051 6098 322, FAX +39 051 6098 084,
piero.desabbata|matteo.busanelli|arianna.brutti|angelo.frascella}@bologna.enea.it,
²Dept. of Computer Science, University of Bologna, Mura Anteo Zamboni 2, Bologna, 40127, Italy
Tel: +39 051 2094871, Fax: + 39 051 2094510, Email: gessa@cs.unibo.it

Abstract: Nowadays enterprises require the development of new solutions that could enhance the interoperability among the partners of a supply chain; a key point in this scenario is represented by data integration mechanisms between such enterprises. In this context semantic descriptions of data and information is emerging as a critical component in a really interoperable framework upon which to implement this integration. Starting from these considerations, the authors propose in this paper an approach to support a document-based framework for a standardisation initiative (based upon the ebXML meta-model and the ISO 11179 compliant vocabulary) introducing ontology-based descriptions in order to exploit semantic information to improve the data integration between enterprises. The use context of this work is the Textile/Clothing sector, where the composition of heterogeneous information systems is really critical.

1. Introduction

The development and the adoption of solutions for e-business can result critical in the scenario of global competition among the enterprises, regardless the production sector. Nearly each commercial activity can exploit ICT to streamline and to optimise its business processes and relationships. Anyway, the system heterogeneity to compose and the differences among the adopted business models could discourage the integration process between the enterprises involved in the supply chain. The effort to tackle the issue takes basically two approaches: the development of new interoperability software architectures, and the raise of international standardisation initiatives. As it could be guessed, these two aspects should be strictly related.

In fact, often standards are too difficult to develop[10] and cumbersome to adopt[12], and need software tools to be managed and used; on the other hand software architectures need common agreements upon which to establish some kinds of communication – basically, a standard. These issues are particularly critical in those sectors where the majority of the enterprises are SMEs (Small Medium Enterprises); in the Textile/Clothing, for example, EDIFACT standard has not represented an effective solution for the sector.

A mechanism to improve standard usability, maintenance and efficiency could reside in the formal description of the semantic of the standards, in order to ease their management and to improve their powerfulness. The semantic description could besides be the bridge to integrated more efficiently the business model views that are local to the enterprises with a
global external view common to the whole business sector. In this perspective semantic descriptions and the bounded technologies are intended as infrastructures that act 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. Especially, this semantic description regards both the general concepts inherent the production processes and, more precisely, the data model implemented by the framework. Our aim is two fold: from one hand, we want to ease the management of the framework, and, on the other hand, we consider this as the first step to improve the interoperability of our framework towards other initiatives, standards or frameworks. To this aim we generate in an automatic manner an ontology starting from a vocabulary of business terms that represents a widespread agreement between the subjects involved in the sector; the ontology is the base upon which to apply semantic integration mechanisms between different systems (see section 2).

More in details, we are now applying our approach in the Textile/Clothing sector. This sector is characterised by the large presence of SME, whose integration is one of the key points for the improvement of the production processes. Our ontology derives from a vocabulary defined by a CEN/ISSS initiative to build an European standard for the sector, and it is strictly correlated with the interoperability framework that is developed.

The paper is organised as follows: first of all, section 2 provides a brief description of the main initiatives in the business scenario to define interoperability frameworks in various business sectors, and mentions some approaches to provide and to exploit semantic descriptions; in section 3 we describe our solution to extend a vocabulary of business terms with a semantic description. Section 4 presents the implementation of our approach; section 5 summarises briefly the use context (within the Moda-ML initiative) of our effort. Section 6 provides some conclusions and future developments of our work.

2. The semantic issue for interoperability framework

The interoperability issues require the definition of proper software architectures and the development of standards that can represent a common base upon which establish more and more efficient business relationships.

The last developments [7][8][9] in this field have adopted basically XML as the basic syntax to exchange structured information, and have besides highlighted the relevance of the semantic aspect in the interoperability problem. The semantic description of a framework is particularly relevant in the context of the document-based interoperability frameworks, that define as basis for the B2B integration a set of complete business documents.

These frameworks in general underpin upon a set of well-defined components used to build business document templates. With this approach, the final document templates are assembled using the proper components chosen among those in the vocabulary; these components represent basically a vocabulary of reusable business terms that are organised and structured depending on the characteristics of the target document templates, usually through a schema language like XML Schema; a list of the most relevant follows:

1. The CommerceNet eCO framework [13] takes an horizontal approach. Regarding the content layer, eCO provides a set of XML core business documents called xCBL – XML Common Business Library. xCBL documents contain general information exchanged during business transactions. This information is common for many different application domains.

2. ebXML[2], that is a horizontal meta-framework to develop specific frameworks; it provides three types of components upon which to structure business documents: core components, domain components and business information objects.

4. UBL [1] is the first implementation of the ebXML Core Components. It provides a narrow set of cross-domain business documents and a more wide set of reusable XML data elements.

In order to efficiently elaborate the instances of the business documents, a deep and exhaustive knowledge of the meaning of the documents is needed. In particular, an automatic elaboration of the documents requires a formal description of the semantic associated with the data and with the terms contained in the documents. Any of the aforementioned frameworks comprises an effort to provide the semantic of the documents.

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; the idea is to express information contained into the (web) documents in a meaningful way accessible to the applications. The Semantic Web thus would result as an infrastructure that provides semantic information about documents, upon which different applications can be implemented; in the e-business context, the semantic 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.

Finally, many research prototypes [6][15][16][17] are now exploiting ontologies and semantic web technologies to improve interoperability. These efforts aim both to provide the semantic description of the frameworks, and to exploit this description for data integration. In order to compose the difference between two heterogeneous systems, these approaches build the semantic models of the systems, and allow data integration performing transformations or mapping operation on the business documents, often adopting different abstraction layers for the target model. What really lacks nowadays is the application of these results to business sectors.

3. A semantic description for a business vocabulary

In the context outlined in the previous section, our approach building an interoperability framework was designed around two main aspects: the definition of a B2B protocol adopted by the framework, and the definition of a set of tools that provides the functionalities needed to exchange business information exploiting the defined protocol.

In order to provide the B2B protocol, we have defined a vocabulary of business terms tailored for the target business sector (our use case regards the Textile/Clothing sector, see section 5), together with a software architecture for the management and maintenance of the framework itself (see [4] for details). The vocabulary has been developed according to the main standardisation initiatives and specifications: in particular we followed the ebXML model[3] 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 structure of the documents (see [5] for details about the approach followed developing the vocabulary). This architecture was originally designed for the automatic generation both of the XML schema templates and their user documentation; we call our architecture the “document factory”.

We have also 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 for the definition of a collaborative framework:
2. Structures of the documents.
3. General concepts treated by the framework.

These domains represent the basic layers needed for an exhaustive description of a framework that aims to give raise to a standardisation initiative for a business sector.

We have now 1) integrated our framework with the semantic description of each of these domains 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. We expect the ontology to support data integration between the enterprises, standards and B2B protocols, to facilitate the framework development and maintenance, and to acts as the interface towards the semantic web.

In the field of ontology development it does not yet exist a well established design approach[18]: it in general results to be an iterative process, where the ontology being developed is continuously evaluated in order to drive the development process itself.

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 the tool that build automatically the whole ontology following the pattern fixed in the draft.

The developed ontology reflects naturally the structure of the vocabulary; for each of the semantic area (domain) listed above, we identified specific basic concepts used as cornerstones to build the final 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 represents also a starting point for the browsing of the ontology. For example, we have fixed the basic concept of Process that has been mapped in the superclass BusinessProcess and in the corresponding sub-ontology BusinessProcess (that is described in the BusinessProcess.owl file). The final ontology is constituted then by all of these subclasses.

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, and eased its maintenance and its future development. The following sections deepen the content and the structure of the semantic areas listed above.

1.1 The process area

This area models the business scenario of the target business sector. It includes the following basic concepts:

- Process: a generic process of the supply chain.
- Activity: a generic activity within a process.
- Document: a generic business document exchanged between two actors; it corresponds to a single transaction within an activity.
- Actor: a business 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 the 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 structures.

3.2 The document area

This semantic area defines the structure of the terms defined by the vocabulary and used within the business documents (in the following we call them components), and consequently reflects also the structure of the business documents themselves. Anyway, the designed ontology does not model the syntax of the documents: this task is achieved through a proper XML Schema for each document (see [5]); we have rather decided to highlight the relationships between XML types, elements and attributes, and the semantic content of the documents, that is strictly related with the business sector the documents are bounded to (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 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.

The XMLElement class has been provided with a set of properties (like “hasAlwaysChild”) used to maintain the relationships (and the hierarchic structure) between the various elements. 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 treated by the framework, relating them with the XML components; this is surely the most interesting area for the semantic description of the business sector.


The three dimensions used to characterise the components of our vocabulary 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 an attribute) used by the documents; the Representation Terms is finally 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 the object “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.

Each object of the ontology has specific properties with which is linked using the relation hasProperty; obviously, each object can have many properties.
Considering the final structure of the resulting global ontology, we highlight as there are some overlaps between the semantic areas: for example, our Document 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 ontologies in a common vision of the framework.

4. The ontology generator

The vocabulary of the business terms is implemented using a database that provides an exhaustive description of each XML component (defining their properties like XML types, length and so on). The database defines also all the relationships and dependences between the documents, the elements and the attributes, maintaining thus an implicit definition of the structure of the business document templates. As we explained in section 3, our vocabulary maintains also the description of the possible business scenarios for the sector.

Once we have defined the structure of the target ontology we aim to produce, we developed a software tool able to generate in an automatic manner the OWL files starting from the information held in the database. This allows us to produce outright a semantic description (the ontology) of our framework each time we update, integrate, or extend it.

The ontology generator consists of a set of software modules. It will be used periodically, basically each time a new version of the framework is realised, and thus works as a batch application that does not require a specific user interface. We just fixed some parameters to drive the generation of the ontology (for example, we can set file names, ontology names, namespeces, name of some main classes and so on.). The ontology generator has been developed upon the Java architecture J2EE [21]: each software module consists of Java classes. This choice is justified by the availability of useful and strong OWL API for this architecture. 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 Protegè.

The development of the application (fig. 1) has involved:

1. the development of the DictionaryConsole Javabeans that extracts data from the database.
2. the development of the OntologyManager Javabeans that builds the ontology and produces the OWL files using the Protégé OWL-API.
3. the development of a web interface using servlet and JSP pages that allows the staff to remotely access to the tool.

Figure 1. The ontology generator software architecture.
A relevant aspect that we have considered in this activity is the interaction of the resulting ontology with other ontologies: it, in fact, is not thought to work in isolation respect other knowledge domains, nor we claim to model everything within our ontology. The main aim that underpins the ontology development is the creation and specification of common models that could guarantee the needed interoperability among heterogeneous systems. In particular 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[19], that defines very general concepts, like time and number.
- The MILO ontology[20], that defines more specialised concepts. It defines for example concepts like product, order or delivery, tailored for the business, but also generic terms that can be used in specific sectors, like fabric, that obviously regards the Textile/Clothing sector.

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

5. The use context

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 issue, also for the large presence of SMEs. This effort originated within the Moda-ML project (www.Moda-ML.org), born out of the collaboration between a number of research organizations, including ENEA, Politecnico di Milano, Gruppo SOI, Domina, Institut Francais Textil Habilliment (IFTH) and a group of leading Italian Textile/Clothing manufacturers. Supported under the IST Programme of the Fifth Framework programme, its results (that include the definition of a vocabulary of business terms and a set of business document templates tailored for the target sector) were included in the final specifications of the European standardisation initiative TexSpin, promoted by Euratex (European industry trade association of T/C industry) and by CEN/ISSS.

After the conclusion of the Moda-ML project, in 2003, the partners have maintained a technical group assuming the denomination of ‘Moda-ML initiative’ and have improved the links with other experiences in Europe (i.e. eTexML, e-Chain, TextileBusiness, T2T); the group is now improving the framework and supporting the industry adoption through the participation to many activities and to the CEN/ISSS TexWeave initiative (www.texweave.org).

The last version of the framework defines a vocabulary composed of 494 terms, a set of 32 business documents, and specifies 8 business processes for the enterprises. The generated ontology includes 373 classes, 44 relationships and 1098 instances.

6. Conclusions

This paper describes an approach to provide a semantic description of a business vocabulary. The main outcomes of our efforts are:

- The definition of an architecture for the management of the semantic of an ISO 11179 compliant business vocabulary.
- The development of a software tool that is able to automatically extract a semantic description from the vocabulary implemented in the database.
Since ISO11179 is the basis of many standardisation initiatives (like UBL) this result could be 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 for the Textile-Clothing sector, named ONTO-MODA, within the European research project LEAPFROG IP (www.leapfrog-eu.org). We consider also these results as a bright premise for further developments in three areas:

- mapping and integration mechanisms to interface each other heterogeneous data models defined by different subjects, like standards and ERP systems;
- studying on the complementary approach of the definition of business vocabularies starting from an ontology;
- linking of the ontology with other (external and/or upper level) ontologies.

References

[1] B. Meadows, "Universal Business language (UBL) 1.0", cd-UBL-1.0, OASIS, September 15 2004; docs.oasis-open.org/ubl/cd-UBL-1.0/