Semantic enrichment of standard-based electronic catalogues
Joao Sarraipa, Carlos Agostinho, Hervé Panetto, Ricardo Goncalves

To cite this version:

HAL Id: hal-00391403
https://hal.archives-ouvertes.fr/hal-00391403
Submitted on 3 Jun 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Semantic Enrichment of Standard-based Electronic Catalogues

João Sarraipa*, Carlos Agostinho*, Hervé Panetto**, Ricardo Jardim-Gonçalves***

* UNINOVA-GRIS, Group for the Research in Interoperability of Systems - Instituto de Desenvolvimento de Novas Tecnologias, Caparica, Portugal (Tel: +351 21948365; e-mail: jss@uninova.pt; ca@ uninova.pt).
** CRAN, Nancy-University, CNRS, FRANCE (e-mail: Herve.Panetto@cran.uhp-nancy.fr)
*** Universidade Nova de Lisboa, Caparica, PORTUGAL (e-mail: rlq@ fct.unl.pt)

Abstract: Today, enterprises are facing serious interoperability problems concerning the exchange of electronic data. Due to the proliferation of terminology, organizations from similar business environments have trouble cooperating, and supply chains are experiencing difficulties exchanging electronically vital information, such as catalogue data. In order to solve this problem, standardization communities are working to define formalized structures for catalogue and product data. However, standards by themselves do not solve semantic interoperability issues. For instance, a group of enterprises which share catalogue information in their business activities need to have a common semantics to understand each other. Otherwise their systems might understand the data structure but not its meaning. This is today a major challenge in modern enterprise integration. This paper contributes to achieving seamless product oriented enterprise interoperability by proposing a framework based on knowledge representation elements to support the semantic enrichment of standard-based electronic catalogues.

Keywords: Semantic Interoperability, Knowledge Representation, Knowledge-based systems, Standards.

1. INTRODUCTION

In the advent of globalization, e-business and electronic data sharing are becoming an integrative part of business networks. However, Interoperability of its systems and applications is an issue when organizations are rethinking their strategies and are looking forward for new business relationships (Sahin and Robinson, 2002). The exchange of information and documents between partners often cannot be executed automatically or in electronic format as desirable, thus causing inefficiencies and cost increase (Brunnermeier and Martin, 1999). This is primarily due to incompatibility problems among the several information representation structures used by the different software applications along supply chains and business networks (Ray and Jones, 2006).

With this diffuse range of systems, industry has had its development of trading partnerships restrained, e.g., inhibiting the shared fabrication of products, software solutions. These barriers are real factors that prevent innovation and development. Therefore, standardization rapidly became an evident priority, and several dedicated reference models covering many industrial areas and related application activities, from design phase to production and commercialization, have been developed enabling industrial sectors to exchange information based on common models (Jardim-Gonçalves et al., 2006a). In that sense, one of the most important sets of standards for representation of product information in industrial environments is ISO 10303, commonly known as the Standard for the Exchange of Product Model Data (STEP) (ISO/TC184-SC4, 2008). It encompasses standards for several industrial sectors as the automotive, aircraft, shipbuilding, furniture, construction, etc.

In the furniture industry, the funStep initiative (funStep, 2008) engaged in STEP to develop a standard for furniture catalogue and interior design representation, formally known as Application Protocol 236, i.e. AP236 (ISO, 2006). However, the use of the AP236 or any other STEP Application Protocol alone does not solve all the interoperability problems. Each stakeholder can have its own nomenclature and associated meaning for their business products. Therefore the information exchanged, in spite of sharing the same structure, still may not be understood by all business partners. This lack of interoperability is related to the semantics of the contents exchanged, and the authors, under the funStep initiative, are proposing the semantic enrichment of the furniture product data as a solution. The main objective is to organize the knowledge associated to the furniture products in order to enable a full understandable business messages and catalogue exchange.

This paper takes the furniture industry example to propose a framework for semantic enrichment of standard-based electronic catalogue data. The framework is built upon different Knowledge Representation Elements (KREs), namely a dictionary, a thesaurus, a reference ontology in the furniture domain and the AP236 standard itself. Together, these KREs establish the syntax and the lexicon to be used in the furniture domain. Each concept has its own definition translated to different languages, and some of the concepts are classified as the reference ones to be used by the community in their information exchange. Specific links between all the KREs enable the establishment of ontology mapping solutions, so that enterprises can keep internally their own terminologies and classification systems, and still be interoperable with their business partners.
2. SEMANTIC INTEROPERABILITY REQUIREMENTS

The ISO 16100-1 (2002) standard defines manufacturing software interoperability as the ‘ability to share and exchange information using common syntax and semantics to meet an application-specific functional relationship through the use of a common interface’. In alignment with this statement, the funStep initiative group intends to facilitate interoperability in the exchange of furniture catalogues and information between stakeholders. The furniture supply chain is characterized mainly by five stakeholders: suppliers, manufacturers, retailers, e-marketplaces and interior designer/architects, whose characteristics and relationships lead to different communications implementation requirements. According to Sendall et al. (2001) the analysis of an use case is one way to verify the system’s functional requirements. Thus, in order to allow a common understanding among developers, system users, and domain experts, the authors are considering an International Product Business (IPB) use case, aiming to represent situations which could facilitate the description of how the knowledge is handled in an international plan of the furniture business. This leads research to complex situations related to semantics interoperability, as for instance multilinguism or regional standards.

In IPB, the globalisation phenomena and the international product business situation drove the authors to the “leather couch” scenario analysis. This scenario is related to a “leather couch” request for quotation where a customer chooses, in a furniture catalogue, a foreign product. The customer starts by asking the supplier for a “leather couch”. The retailer shows the customer a set of catalogues with various types of “sofas”. The customer looks at the catalogue and chooses the model “XPTO”. However, in the catalogue it is not described if such model is available in “leather”. Thus, the retailer has to ask manufacturer of the chosen model if it can be made in “leather”. Since the model’s manufacturer is in French, the request has to be translated. The retailer sends a fax with a request for quotation of a “XPTO en cuir” (translation of “in leather” to French). The retailer asks the customer for his contacts (e.g. email; phone; etc.) in order to be in touch.

The manufacturer does not have “leather” in stock, he has to contact his supplier. Since his “leather” supplier is Spanish, he translated the request to “cuero” (translation of “leather” to Spanish). The supplier answered his request. Nevertheless, the manufacturer did not believe in the answer because the product description said “piel” (Spanish synonym word for “cuero”) instead of “cuero”. When the Spanish supplier received the reply asking for confirmation, he laughs because of the supposed misunderstanding that the French thought was happening. In order to avoid further interaction problems, the Spanish answered again using the “right” word – “cuero”. Once the manufacturer received the quotation, he replied with the quotation for producing a “leather XPTO” sofa. After four days the customer received the answer.

The funStep initiative already defined a standard, i.e. AP236, for the exchange of data that could be used in this use case scenario communications. Nevertheless despite the information exchange with AP236, semantic interoperability could still be improved. From a first analysis to the presented use case scenario, it was emphasized the following set of statements that describe the necessity of having a structured furniture knowledge organization:

1) The retailer knows that “couch” is equivalent to “sofa”, and the catalogues containing these products mostly designate them as “sofas”. Thus, the reference word is “sofa” instead of “couch”.
2) The retailer only shows to the customer the catalogues that have sofa models. This implies a catalogue selection based in a classification item – “sofas”.
3) Once the customer wants the model “XPTO” in “leather”, the retailer had to translate this request to the language of the chosen model’s manufacturer.
4) In the interaction between the manufacturer and the supplier, it was identified the need of having defined reference terms and concepts for each translated language.
5) Since the retailer is English and the manufacturer is French, the quotation was converted to the right currency.

3. FUNSTEP KNOWLEDGE ARCHITECTURE

Nonaka et al. (2001) defines two kinds of knowledge: 1) Tacit knowledge: that people carry in their minds, which provides context for people, places, ideas, and experiences; 2) Explicit knowledge: that has been or can be articulated, codified, and stored in certain media.

Fig. 1. funStep knowledge architecture and domain intersection

funStep endeavours to gather the tacit knowledge that furniture domain stakeholders hold into machine interpretable knowledge bases. Closing this gap, funStep will obtain the (explicit) knowledge which should be stored in a structured organized way, where syntax and lexical semantics are integrated (Pustejovsky, 1993). For reaching that objective, the authors are proposing to integrate the funStep standard (AP236) with the reference funStep Lexicon, which embodies the reference concepts and semantics, and with a funStep ontology, which embraces product classification to its related properties. This leads to the knowledge architecture definition (Fig. 1). Therefore, the integrated knowledge is composed by four KREs: the funStep Ontology; the funStep
Thesaurus; the funStep Dictionary, and the funStep AP236 ISO Standard.

For a good explicit knowledge representation, it is needed to have significant input from the tacit source (i.e., domain experts). Thus, such characteristic requires a knowledge architecture enabling the management of the evolution between the KREs. The evolution of the first three KREs leads to the funStep Lexicon establishment which is an abstract KRE in the sense that it is composed by thesaurus contents, i.e. concepts and definitions. On the other hand, the funStep explicit knowledge KRE is another abstract KRE since it is composed by the addition of the funStep Lexicon with the ontology and the standard itself. In conclusion the funStep explicit knowledge represents all the furniture machine interpretable knowledge where the funStep dictionary and the thesaurus are supporting KREs to the funStep Lexicon establishment and maintenance.

Each one of these KREs has a particular role and different objectives in the overall funStep Product Knowledge Architecture, where their focus domains intersect each other partially (top right part of Fig. 1). The funStep thesaurus domain is totally included in the funStep ontology, while the other KREs have particular information which is not shared with the others.

### 3.1 The funStep Dictionary

A domain dictionary has been found to be one of the most useful tools for a domain analysis. The dictionary lesson a great deal of miscommunication by providing users with information: 1) in a central location to look for terms and abbreviations that are completely new; 2) where definitions of terms are used differently or in a very specific way within the domain (CMU, 2007).

**Sofa:** Comfortable seat with raised arms and back filled or covered with soft material and long enough for two or more people to sit on.

*Related words:* Couch

*Translation:* Fr. Canapé; Port.Eu Sofá; Esp. Sofá

Fig. 2. The funStep Dictionary

The funStep dictionary supports a multilingual collection of terms, thus enabling a correct coordination between international partners. Also, the terms are associated to a description, synonyms, other related terms, and optionally multimedia items, such as images, sounds or videos. Continuing with the “leather couch” scenario, Fig. 2 represents a view of the dictionary with the meaning of the term “Sofa”; its picture; its related words – in this case it is “couch”; and some translations (e.g. “canapé” in French; “Sofá” in Spanish and Portuguese).

### 3.2 The funStep Ontology

An ontology is an explicit specification of a conceptualization (Gruber, 1993). Thus, ontologies could provide a basis for expressing and structuring the knowledge of an organisation. In the context of the funStep initiative, its main objective is to represent all the knowledge associated to furniture products, enabling semi-automatic classification.

The funStep ontology is being developed, initially based simply on search criteria and identification which were used for furniture product classification in electronic commerce. Nowadays, we are using semantic comparisons, basic lexicon establishment, harmonization among other ontologies and other operations on knowledge base representations. Fig. 3 depicts a subset of the funStep ontology taxonomy emphasizing the furniture product characteristics mentioned in the “leather couch” scenario – classification as “leather sofa”.

![Fig. 3. Subset of the funStep Ontology taxonomy](image)

### 3.3 The funStep Thesaurus

The basic lexicon establishment is reached by the development of a thesaurus on the domain – the funStep Thesaurus. This is composed by a set of domain reference terms and concepts, clustered on the basis of their similarity, organized by means of semantic relationships (e.g., equivalence, subsumption, generalization, disjunction ...), thus enabling a better retrieval process of semantically related terms (Missikoff et al. 2004). A thesaurus can serve as a controlled vocabulary where terms are constrained to its domain-specific meanings, avoiding the problem of ambiguity (Gatlin, 2005). The funStep thesaurus envisages a multi-national scope of vocabulary, where terms with the same meaning coexist in multiple languages. Multilingual thesauri can be used to translate queries, by expanding the query to one or more target languages (Ballesteros and Croft, 1996). Still in the “leather couch” use case scenario, the multilingual thesaurus addresses the definition of the related words “Leather”, “Cuir”, and “Cuero”.

### 3.4 The funStep AP236 Standard

To allow enterprise applications to interoperate seamlessly in information exchange, there is a need for a unified and standardized representation of product data (Ray and Jones, 2006; Jardim-Goncalves et al., 2007), i.e., the funStep ISO AP236 standard (ISO, 2006). AP236 defines a reference open structure for catalogue and product data representation under industrial domains of the furniture sector, helping on the information interoperability at a syntactical level. The External Classification; and Multilinguism modules of funstep standard are examples of relevance for semantic enrichment.

A) External Classification

For an improved business, networks of organizations may define, or use shared reference ontologies or thesaurus, instead of legacy taxonomies. In this case, when exchanging
product information, they should classify their products using that reference nomenclature. AP236 provides a mechanism for that, i.e. the external classification.

External classification enables a direct link between the context and the classes of products. Nevertheless, in this case, the mentioned link uses meanings that are expressed in libraries physically described in remote locations. Fig. 4 illustrates how external classification works. The manufacturer links with the funStep reference ontology, instead of sending its own, using the AP236 entities: “external class” and “external class library”, to identify the externally defined concepts, and “classification assignment” to establish the link with the product in the catalogue.

Fig. 4. AP236 External Classification Example

B) Multilinguism Issues

For an internationalization of the market, companies must be able to send their products and associated information in several languages, especially the official language that the receiver company speaks. The multilinguism resource of AP236 addresses such need, so that any organization could receive data in their native language independently of where the receiver company speaks. The multilinguism resource of AP236 addresses such need, so that any organization could receive data in their native language independently of where the receiver company speaks. Therefore, multilingualism is the act or process of adding specific meaning elements to some knowledge representation structure in a domain, to help on the information clarification (Boudjilida and Panetto, 2007). The funStep standard semantic enrichment is characterized as being performed at two basic levels (Boudjilida, et al, 2007): 1) Terminological annotation, by using the terms identified in the Lexicon (Thesaurus); 2) Semantic annotation, by using concepts and expressions drawn from the Ontology.

Fig. 5. AP236 Multilinguism Example

Multilinguism allows the translation of the attributes of AP236 entities in any number of languages. In Fig. 5, it is possible to notice that the AP236 entity “language” is used to indicate the reference language. The entity “language_indication” points to the attribute that one wants to translate, and the “attribute_translation_assignment” specifies the translation itself. In this case, it is the value of the attribute “name” that is being translated from the French “cuir” to the English “leather”.

4. THE FUNSTEP SEMANTIC ENRICHMENT FRAMEWORK

Semantics is the study of language units meaning and their combinations (Willerval et al., 1989). Therefore, semantic enrichment is the act or process of adding specific meaning elements to some knowledge representation structure in a domain, to help on the information clarification (Boudjilida and Panetto, 2007). The funStep standard semantic enrichment is characterized as being performed at two basic levels (Boudjilida, et al, 2007): 1) Terminological annotation, by using the terms identified in the Lexicon (Thesaurus); 2) Semantic annotation, by using concepts and expressions drawn from the Ontology.

Fig. 6. funStep Knowledge Framework

Both levels are conducted through a knowledge framework based on the KREs identified: the funStep dictionary, thesaurus, ontology and standard data. Matching the requirements identified in the use case presented, the author’s proposal for the funStep framework (Fig. 6) is built up on web services technology enabling interoperable open services over the internet between the funStep knowledge client systems and the funStep knowledge server (Jardim-Gonçalves et al. 2006b). This architecture is composed by three parts: 1) natural language dictionary server; 2) funStep knowledge server and 3) funStep knowledge clients.

4.1 Natural Language Dictionary (NLD) server

The NLD server is composed by an external component to the funStep initiative, having as its main objective to complement the domain dictionary in the translation procedure of natural language information used within the AP236 standard data. For instance, the request for quotation sentence sent in the use case scenario was translated from “XPTO in leather” to the French “XPTO en cuir”. The
"leather" word was translated by the domain dictionary, but the “in” word it was translated through a NLD.

4.2 The funStep knowledge server

The “funStep” knowledge server” is composed by the four funStep KREs described in section 3, whose relationships enable the semantic enrichment of the standards data.

The AP236 standard data has links to the ontology for products classification, and to thesaurus concepts for implicit annotation to its data representation elements. This is made as explained before for the External Classification example. Thesaurus concepts and dictionaries terms (from domain and natural language dictionaries) are used in the knowledge server data representation and multilingualism translations. Some of the terms and meanings in the domain dictionary are related to thesaurus elements, which facilitate the retrieval of the related reference concepts, e.g. the “couch” concept was replaced by the reference concept “sofa”.

The ontology is closely related to the thesaurus, sharing a similar structure. This enables the usage of concepts definitions from both KREs for semantic clarification and to establish a relationship between them. Therefore, the funStep Ontology could be easily translated taking as source the thesaurus for the target language.

4.3 The funStep knowledge clients

The “funStep” knowledge clients” are interfacing the framework engine with the funStep knowledge users, which can be administrators, customers, or stakeholders.

The role of funStep system administrators is mainly related to systems evolution and maintenance, i.e., KREs in this framework case.

Traditionally, in the furniture sector, customer interfaces are mainly focused on simple product search. This framework intends to go further ahead, enabling software developers with skills to develop enhanced intelligent products search engines based on knowledge reasoning.

The stakeholders’ interfaces are related to standard semantic enrichment itself. The terminological and semantic annotation procedures follow the external classification and the multilingualism mechanisms previously explained. Nevertheless, because of the normal reluctance of SMEs in adhering to a reference knowledge architecture, the need for a semantic alignment feature has been identified. It intends to facilitate SMEs in becoming a part of a structured knowledge organization without needing to change their own knowledge lexicon and classification structures. This could be possible through semantic mappings between the involved KREs.

Semantic mapping is an activity that attempts to relate the vocabulary between two organizations that share the same domain of discourse. Sarraipa et al. (2008) defined a methodology for enterprise reference ontology development, which uses a mediator ontology able to represent semantic operations: semantic mismatches; semantic transformations; mappings; and other (e.g. versioning). Such mediator ontology should then be adopted, as KREs mappings facilitator.

The transformation resulted from the mapping of this example can be recorded in the mediator ontology as an axiom, which, in this specific case, is based on the following equations:

\[ FR(x) = n\_seats(x) \]  
\[ FE(y; z = \text{dimension}(\text{size}(y), \text{depth}(z))) \]  
\[ FR(x) = FE(y; z) \]  
\[ n\_seats(3) = \text{dimension}(\text{size}(215), \text{depth}(90)) \]  
\[ y = 50 + 55x \land z = 90 \]  
\[ FR(x) = FE(y; x) \]  
\[ n\_seats(x) = \text{dimension}(\text{size}(50 + 55x), \text{depth}(90)) \]

Equation (1) indicates a function that represents the “Leather couch” property, which, in this case, is related to more than one property in relationship to the manufacturer representation (2). In order to define the transformation which relates both representations, it is stated an equality between both expressions (3). After analysing empirically all the existent values that these expressions could take ((4) shows one case), it was defined two linear equations (5) which relates (1) and (2). At the end it was reached an expression that establishes a semantic relationship between both representations and that establishes the transformation equations related to each variable (6) and (7).

Thus, this mediator ontology can be used for semantic translations between enterprises exchanging data, which do not share the same semantics as a funStep knowledge organization.

5. CONCLUSIONS
The advantage when adopting the proposed knowledge framework is to enable enterprise systems with semantically seamless communication with other actors using the reference knowledge. It allows enterprises to keep their internal terminologies and classification systems, and still remain interoperable with their business partners, through the usage of knowledge mapping procedures. The funStep knowledge framework uses different KREs as catalysts to enable such semantic interoperability. Together, domain dictionary, thesaurus, reference ontology and the AP236 standard itself act as explicit knowledge repository and reference lexicon for the application domain. To complement, the domain dictionary in the translation procedure, a natural language dictionary is used. Finally, the framework also considers the interaction with the users, i.e. administrators, customers, and stakeholders, using web services which not require special tools or knowledge to implement instead of what happened with traditional Enterprise Application Integration systems.

In the past, the funStep initiative identified two major levels of compliance in the furniture communications: (level 1) - not funStep compliant – where the messages exchanged are following any kind of format other than the AP236 standard; (level 2) - funStep compliant – where, the exchanged messages are compliant with the AP236 standard. Inside this level, there can still be different sub-levels according to the parts of AP236 implemented. As a major contribution, the work proposed in this paper is defining a funStep Knowledge framework that adds a kind a semantics compliance, i.e. (level 3) - funStep knowledge compliant. In this type of compliance, systems communications would be syntactically compliant with the AP236 standard and semantically compliant with the reference funStep knowledge.

REFERENCES


