Modeling and Exchange of Product Classification Systems using XML

Joerg Leukel, Volker Schmitz, Frank-Dieter Dorloff
Department of Information Systems, University of Essen, Germany
{joerg.leukel | volker.schmitz | dorloff}@uni-essen.de

Abstract

Product classification systems play a major role in searching and comparing offered products on electronic markets. Especially in case of large multi-vendor product catalogs classified data becomes an important asset and success factor. The most known systems are UNSPSC and eCl@ss, however they are still developing, and new systems are emerging as well. Classification systems differ not only in content but also in structure from each other. The management and exchange of the systems between market partners must be able to get along with these differences. A common structure model describing classification systems is missing so far. This paper discusses the design of classification systems and argues to develop standardized messages using XML Schema for the transmission of classification systems.

1. Introduction

The task of product classification is to assign each product to a product group corresponding to common attributes or application areas. Though classification systems are not a new phenomenon of B2B e-commerce; they are already in use as an instrument of structuring since decades. The usage fields are very broad. They extend from manufacturing, costing and sale (e.g. product catalogs) up to national and international economic statistics.

In B2B e-commerce classification systems gain a new meaning and function. They are an instrument for the access to large e-catalogs. Standardized and supplier-independent classification is an elementary requirement for efficient product search and qualified comparison of products in electronic markets and other catalog-based procurement systems [1]. To describe products in a uniform manner, some classification systems define so-called sets of attributes. A set of attributes is assigned to a classification group and contains the necessary product attributes. In e-catalogs that claim to support the classification system each product has be described by the group-depend set of attributes.

Just as little as there is a generally accepted XML standard for business documents today, we cannot expect that a single classification system will prevail worldwide and for all branches of industry. Rather we see the development of vertical classification systems along the requirements of branches and markets. In consequence classification systems become an object of data management and data exchange [2]. For the efficient handling of classification systems a model that describes classification systems is missing so far. Catalog applications implement search and navigation mechanisms on basis of classification systems. However they cannot import the classification system definitions in a standardized format.

2. Paper Organization and Related Work

Goal of this paper is to identify the requirements on modeling classification systems. First of all, we will relate this task to other research work. The main task is an empirical analysis of both classification systems and XML catalog standards. The result is a comprehensive set of design parameters describing the structure of classification systems. We will use this set for the analysis of selected classification systems (Section 4). Afterwards we will ask in Section 5 how XML catalog standards already support the transmission of classification systems. Finally, we will draw conclusions and mention further work to be done.

Research literature show two main working areas regarding classification systems. The first area deals with the application of classification systems and the requirements on them (e.g. [3]). Structural and content wise design aspects are described in [4]; they should serve for the evaluation and development of new systems. The structural aspects are derived on the basis of three systems. In [5] the problem of different schemas for the categorization of products and their attributes is examined from the view of relational databases. The conclusion is that concepts of schema integration cannot be transferred directly.

Approaches for the integration of classification systems on a semantic level form a second area. Classification systems are seen as product ontologies that enable a common and accepted communication in the respective product domain [6]. The specification of ontologies takes place using concepts and languages of knowledge representation. In [7] an integration approach is presented, which imports classification system data using a Wrapper. The supply of this data in a standardized format is not demanded. In [8] a similar, integration-oriented path is

Table 1. Analysis of selected Classification Systems

Area	Design Parameter	eCl@ss 4.1	ETIM 1.1	RNTD 1.4	EGAS 1.0
Attributes	Description	+	-	+	+
	URL	-	-	+	-
	Unit	+	+	+	-
	Synonyms	+	-	+	-
	Groups	-	-	-	-
	Datatype	+	+	+	+
	Value Orders	-	+	-	-
	Unit	+	+	+	-
	Symbol	+	-	+	-
	Mandatory vs. Optional	-	-	-	-
	Attribute Orders	-	+	-	-
	Attribute Inheritance	-	-	-	-
Groups	Description	+	-	+	-
	URL	-	-	+	-
	Synonyms	+	+	+	-
	Multimedia Objects	-	-	-	-
	Set of Attributes	+	+	+	+

taken. [9] proposes an information retrieval approach. The import process of UNSPSC data is not described.

3. Analysis of Classification Systems

The result of our empirical analysis of classification systems and XML catalog standards is a set of design parameters. These parameters are divided into the areas attributes and classification groups. Table 1 shows, which design parameters are implemented in four selected product classification systems: eCl@ss [10], ETIM [11], RosettaNet Technical Dictionary (RNTD) [12] and EGAS [13] that adds sets of attributes to UNSPSC [14]. ETIM and RNTD are vertical systems developed for the wholesale of electro technical products respectively for electronic and IT components.

The systems itself are documented by non-formal and formal specifications. Though only RNTD is specified by an XML-document. All other systems use simple Excel or comma-separated value (CSV) files as containers and provide very few semantics. Since the files differ completely in structure, importing them into target systems is a time-consuming task. ETIM provides an MS Access database file, which is described by an ER schema.

Table 2. Analysis of selected XML Catalog Standards

Area	Content	BMEcat 1.2	OAGIS 7.2.1	xCBL 3.5
Attributes	Description	+	+	-
	URL	-	-	-
	Unit	+	+	+
	Synonyms	-	-	-
	Groups	-	+	-
	Data Type	+	+	+
	Value Orders	-	+	-
	Unit	+	+	+
	Symbol	-	-	-
	Mandatory vs. Optional	+	+	+
	Attribute Orders	+	+	-
	Attribute Inheritance	-	-	+
Groups	Description	+	+	+
	URL	-	-	-
	Synonyms	+	-	-
	Multimedia Objects	-	+	-
	Set of Attributes	+	+	+

4. Analysis of XML Catalog Standards

Before we can analyze XML E-Business standards we must ask, which standards are capable of transmitting classification systems. We observe that many catalog standards are confined to the classification of products by giving a reference to the classes and attributes. cXML, eCX and EAN.UCC belong to this group of not relevant standards. ebXML is a framework and does not specify business documents. In contrast the following standards provide special document or data elements definitions for classification systems: BMEcat [15], OAGIS [16] and xCBL [17]. The analysis was done on data element level. For each standard was tested whether the design parameters can be represented and if so by which data elements. A summary is shown in table 2.

5. Conclusions and Further Work

In this paper we have presented the design parameters of product classification systems in B2B e-commerce. The developed set of design parameters serves as a framework for the analysis of classification systems and XML catalog standards. Latter can be evaluated how they are able to transfer classification systems.

All things considered, none of the four selected industrial classification systems realizes all design parameters. The systems themselves are documented quite differently. The system specifications are often provided in proprietary formats; hence their processing in catalog systems is less automated. Especially, the organizations that develop and maintain classification systems provide no XML data (exception: RosettaNet).

The application of XML e-business standards for the transmission of classification systems is hardly possible, because no one of the analyzed standards is capable to transfer all systems completely. The loss of structural information is in many cases very high. Matching the tables 1 and 2 can proof this.

To solve the described problems we will develop an XML model that covers all design parameters and is able to describe all classification systems. The benefit using XML Schema language [18] instead of ERM, UML or RDF is providing a format immediately, which can transfer real classification systems in all details.

6. References

[1] M. Ginsburg, J. Gebauer and A. Segev, "Multi-Vendor Electronic Catalogs to Support Procurement: Current Practice and Future Directions", *Proceedings of 12th International Bled Electronic Commerce Conference*, Bled, Slovenia, 1999.

[2] M. Stonebraker and J.M. Hellerstein, "Content Integration for E-Business", *Proceedings of ACM SIGMOD/PODS 2001*, Santa Barbara, California, 2001.

[3] A.M. Fairchild and B. Vuyst, "Coding Standards benefiting Product and Service Information in E-Commerce", *Proceedings of ICEIS 2001*, Setúbal, Portugal, 2001, pp. 991-995.

[4] B. Otto and H. Beckmann, "Klassifizierung und Austausch von Produktdaten auf elektronischen Markt-plaetzen", *Wirtschaftsinformatik*, 2001, pp. 351-362.

[5] W. Ng, G. Yan and E.-P. Lim, "Heterogeneous Product Description in Electronic Commerce", *ACM SIGeCom Exchanges*, 2000, pp. 7-13.

[6] D. Fensel, Y. Ding, B. Omelayenko et al., "Product Data Integration in B2B E-commerce", *IEEE Intelligent Systems*, 2001, pp. 54-59.

[7] O. Corcho and A. Gómez-Pérez, "Solving Integration Problems of E-Commerce Standards and Initiatives through Ontological Mappings" *Proceedings of IJCAI 2001 Workshop on E-Business & the Intelligent Web*, Seattle, USA, 2001.

[8] C. Quix, M. Schoop and M. Jeusfeld, "Business Data Management for Business-to-Business Electronic Commerce", *SIGMOD Record*, 2002, pp. 49-54.

[9] Y. Ding, M. Korotkiy, B. Omelayenko et al., "GoldenBullet: Automated Classification of Product Data in E-commerce", *Proceedings of BIS 2002*, Posen, Poland, 2002.

[10] http://www.eclass-online.com.

[11] http://www.etim.de.

[12] http://www.rosettanet.org.

[13] http://www.eccma.org/ega.

[14] http://www.un-spsc.net.

[15] http://www.bmecat.org.

[16] http://www.openapplications.org.

[17] http://www.xcbl.org.

[18] http://www.w3.org/TR/xmlschema-0, 2001.