Challenges in Content Management for B2B Electronic Commerce

Dieter Fensel
Vrije Universiteit Amsterdam
dieter@cs.vu.nl

1 Introduction

The World Wide Web (WWW) has drastically changed the on-line availability of information and the amount of electronically exchanged information. Meanwhile the computer has mutated from a device for computation into an entrance portal of large information volumes, communication, and business transactions. Also, it starts to change the commercial relationships between suppliers and customers. The U.S. Department of Commerce estimates for the business-to-consumer (B2C) area between $4 billion to $14 billion on-line sales in the US for 2000 which is around 1% of the overall sales figures. This is still a small fraction of the overall business figures but its fast growth looks very likely given the fact that the number of internet users grew from 100 to 300 millions between 1997 and 2000. Similar estimates are done for the business-to-business (B2B) area. Forecasts for the dollar value of B2B EC in the US range between $600 billion to $2.8 trillion for 2003. Currently, a large fraction of the B2B transactions are still realised by traditional non-Internet networks, such as those conducted over EDI systems. In this traditional paradigm, direct 1-1 connections and mappings are programmed based on standards like EDIFACT. However, this traditional paradigm does not at all employ the full power of electronic commerce and it is quite likely that it will soon out-ranged by more timely, internet and web-based transaction types.

Internet-based electronic commerce provides a much higher level of flexibility and openness that will help to optimise business relationships. B2B market places provide completely new possibilities for their clients.

- Instead of implementing one link to each supplier, a supplier is linked to a large number of potential customers when linked to the market place.
- A supplier or customer can choose between a large number of potential customers and can optimise their business relationships.

In a nutshell, B2B market places are a middleware, that help their customers to contact a large number of potential clients without running into the problem of implementing a large number of communication channels. However, preventing their customers from the bottleneck of facing exponential growth in the number of implemented business connections faces B2B market places with serious problems.

B2B market places have to deal with the problems of heterogeneity in catalogue, product, and document description standards of their suppliers. Effective and efficient management of different description styles become therefore a key task for these market places. Such a successful content management for B2B electronic commerce has to deal with various aspects like: information extraction from rough sources, information classification to make product data maintainable and accessible, reclassification of product data, information personalization, and mappings between different information presentations. All of these sub-tasks are hampered by the lack of proper standards (or in other words by the inflation and non-consistency of arising pseudo-standards). The talk will focus on these challenges for content management and will discuss potential solution paths.

2 The Sub-Tasks of Content Management in B2B Electronic Commerce

B2B market places need to be open for a large numbers of suppliers and buyers. Their success is closely related to their ability to mediate a large number of business transactions. They are an intermediate layer for business communications providing one serious advantage to their clients. They can communicate with a large number of customers based on one communication channel to the market place. The market places reduce the number of mappings to their user community from n*m to n+m. However, in order to provide this service, they have to solve the significant mapping and normalisation problem for their clients. A successful market place has to deal with various aspects. It has to integrate with various hardware and software platforms and has to provide a common protocol for information exchange. However, the real problem is the heterogeneity and openness of the exchanged content. In consequence, content management is the real challenge in successful B2B electronic commerce. Content management in B2B electronic commerce faces a number of serious problems: Product descriptions are unstructured, they are unclassified, they
must be classified and described in various dimensions because no standard product classifications exist, and product descriptions must be personalised to enable customers to find the products they are looking for.

2.1. Product descriptions must be structured

Suppliers have product catalogues tailored for their potential clients. This information should be made on-line available by a B2B market place. One could think that this may be a simple task because most product catalogues already exist electronically. However, these product catalogues are designed for the human reader. Extracting the actual product information and storing it in a structured format is therefore mainly a manual task. A content management solution provider like Content Europe has more than hundred employees working in so-called content factories to manually structure the product information. In the worst case, they take printed copies of the product catalogues as input. In general, this process has two sub-tasks: defining the product categories and their attributes (i.e., defining the schema), and extracting the actual values for the defined attributes.

2.2. Product descriptions must be classified

After the process of structuring product data each product corresponds to an entry in a table where the columns reflect the different attributes of a product. Similar products are grouped together in the same table. Each supplier uses different structures and vocabularies to describe his products. This may not cause a problem for a 1-1 relationship where the buyer may get used to the private terminology of his supplier. B2B market places that enable n-m commerce cannot rely on such an assumption. They must classify all products according to a standard classification schema in order to exchange information. UN/SPSC is a hierarchical classification, with five levels. Each level contains a two-character numerical value and a textual description of the product category.

Again it is a difficult and mainly manual task to classify the products according to a classification schema like UN/SPSC. It requires domain expertise and knowledge about the product domain. Therefore this process is costly, however, a high quality is important to ensure maintainability and visibility of product information.

2.3. Product descriptions must be re-classified and described in various dimensions

Bottlenecks in exchanging information have led to a plethora of different standards that should improve the situation. However, usually there are two problems. First, there are too many “standards”, i.e., none of them is an actual standard. Second, standards mostly lack important features for various application problems. Not surprisingly, both problems appear also in B2B electronic commerce. UN/SPSC is only one classification schema and it also has serious shortcomings:

- It is not descriptive, i.e., it does not define any attributes for describing the products.
- It is not very intuitive, i.e., neither supplier nor buyer find their products in the classification.
- It is very shallow, i.e., it does not provide enough distinctions for a vertical market place that provides a large number of products from a certain domain.

UN/SPSC is a typical example for a horizontal standard that covers all possible product domain, however, is not very detailed in any domain. An example for another approach is ecl@ss. ecl@ss aims like UN/SPSC to be a horizontal standard for all products and it enriches the pure classification schema by attributes. Another example is the Universal Content Extended Classification (UCEC). It takes UN/SPSC as a starting point and refines it by attributes. Finally, we would like to mention Rosetta Net as an example for a vertical standard describing products of the hardware and software industry in detail. Vertical standards describe a certain product domain in more detail than common horizontal ones. Finally, we have to mention that there are much more examples for such standards [1].

Because different customers will make use of different classification schemas the product information must be classified and described according to several schemas. This objective defines three sub-tasks that have to be performed for successful content management:

- We need to define links between different classification schemas that relate the various concepts and attributes.
- We need to actually re-classify a product. Because there does not need to be a 1-1 correspondence between concepts in different classification schemas, we often require the actual product information to decide about the new classification of it.
- We have to transfer the original descriptions into the new description style.

Each of these sub-tasks is far from being trivial.
2.4. Product descriptions must be personalised

Personalisation is partially similar to re-classification. The buyer wants to have his own personal view on the products, i.e., he requires their re-classification according to his personal taste. However, the difference is that he may not provide an explicit new classification schema. Take a secretary as an example. She may not want to see the full-fledged product catalogue but only the MRO goods that are relevant to her office environment. She may also want to find products classified according to their various business needs or business processes and not according to UN/SPSC where she will never find her eraser. In a nutshell, this task is about defining a view of information based on role descriptions of the different users.

Establishing on-line stores is a routine activity in the meantime. However, most on-line store do not employ the full power of the new medium: Adaptability and intelligence. Shops must be adaptable to user preferences and his specific current context. Then they can fully employ the superiority of the on-line media. Physically rearranging the presentation of the entire offer of a supermarket for each client and each point in time he visits it is not affordable. In case of on-line stores, precisely this can be achieved. Based on the virtual character of the on-line store it is possible to adapt the product presentation to individual customers and customer groups (user profiles and corporative filtering). Personalization brings electronic commerce to its full potential.

2.5. There are even more issues involved

The number of sub tasks we discussed up to now do not provide the complete picture. There are at least two more important sub tasks we have not discussed yet. One is concerned with enrichment of product descriptions. The second one is concerned with the enrichment of product standards. Both are largely complementary.

? In the first case, a structured product description turns out to be incomplete or non-standard according to the standard set of attributes that the classification schema may assume. Then a loop back in the information extraction step is necessary to acquire additional information on the product. From a practical point of view, this process appears often and requires significant resources in content management.

? In the second case, a product standard turns out to be incomplete or not suitable for describing the products in an appropriate manner. Then, life for content manager starts to become hard. They need to play an active role in complex and ill-designed standardisation bodies in trying to overcome many of the obvious bottlenecks in current B2B standards.

3 Conclusions

Content management is the key issue in B2B electronic commerce. In the talk, we will discuss the main obstacles for successful content management, we will identify the main tasks, and we will discuss possible support in mechanising them. Tools based on information retrieval, machine learning, and ontology technology can significantly help to improve efficiency and effectiveness of the overall process.

4 References