

Automated Classification of Product Descriptions

Y. Ding and D. Fensel

Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, NL

Internet and Web technology starts to penetrate many aspects of our daily life. Its importance as a medium for business transactions will grow exponentially during the next years. B2B market places provide new kinds of services to their clients. Simple 1-1 connections are getting replaced by n-m relationships between customers and vendors. However, this new flexibility in electronic trading also generates serious challenges for the parties that want to realize it. The main problem here is caused by the heterogeneity of information descriptions used by vendors and customers. Intelligent solutions that help to mechanize the process of structuring, classifying, aligning, and personalizing are a key requisite for successfully overcoming the current bottlenecks of B2B electronic commerce.

Finding the right place for a product description in a standard classification system such as UNSPSC is not at all a trivial task. Each product must be mapped to the corresponding product category in UNSPSC to create the product catalog. Product classification schemes contain huge number of categories with far from sufficient definitions (e.g. over 15,000 classes for UNSPSC) and millions of products must be classified according to them. This requires tremendous labor effort and the product classification stage takes altogether up to 25% of the time spent for content management.

GoldenBullet is a software environment targeted to support product classification according to certain content standards. It is currently designed to automatically classify the products, based on their original descriptions and existent classifications standards (such as UNSPSC). It integrates different classification algorithms from the information retrieval and machine learning areas and some natural language processing techniques to pre-process data and index UNSPSC so as to improve the classification accuracy. The system helps to mechanize an important and labor-intensive task of content management for B2B E-commerce.