Automated Product Taxonomy Mapping Using SCHEMA

Steven S. Aanen, Lennart J. Nederstigt, Damir Vandić, and Flavius Frăsincar
{steve, lennart}@student.eur.nl, {vandic, frasincar}@ese.eur.nl

In recent years the Web has increased dramatically in both size and range, playing an increasingly important role in our society and world economy. This tremendous growth also means that it is becoming increasingly difficult for a user to find the desired information. Traditional keyword-based search cannot properly filter out irrelevant Web content, leaving to the user the task to select relevant information from search results.

Search failures are manifested in the domain of e-commerce as well. Due to the absence of Web-wide faceted product search, it is difficult to find the product that best satisfies user's needs. Users often switch between Web-wide keyword-based search results and price comparison sites to find the 'best' product. A solution to this problem is to aggregate data from different Web stores and present it to the user in a uniform manner. One way to achieve this aggregation is to use the existing hierarchical product category structure, i.e., the product taxonomy. By matching the product taxonomies from different Web stores, it becomes easier to compare their products. This should contribute towards alleviating the search problems encountered by users when shopping online.

We propose the *Semantic Category Hierarchy for E-commerce Mapping Algorithm* (SCHEMA), which can be used for mapping heterogeneous product taxonomies originating from multiple sources. The algorithm employs *word sense disambiguation* techniques, using WordNet, to find synonyms of the correct sense for the source category name. Furthermore, it uses lexical similarity measures, such as the *Levenshtein distance* to determine the candidate target categories. The best candidate target category is determined using the *Damerau-Levenshtein distance* and a node-dissimilarity penalty. In order to evaluate SCHEMA, its performance is compared on recall and precision with PROMPT [2] and the algorithm proposed by Park & Kim [3].

In his talk, Damir Vandic, PhD student at the Erasmus University Rotterdam, will discuss the results of this study. First, Damir Vandic will present the three main processes of SCHEMA: source category disambiguation, candidate target selection, and candidate target path comparison. Then, the results of the evaluation will be presented, where SCHEMA is compared favorably with two other state-of-the-art mapping algorithms. The work presented here follows from a recent publication [1].

References

[1] S. Aanen, L. Nederstigt, D. Vandić, and F. Frăsincar. Schema - an algorithm for automated product taxonomy mapping in e-commerce. In *Proceedings of the 9th Extended Semantic Web Conference (ESWC 2012)*, pages 300–314. Springer-Verlag, 2012.

[2] Natalya F. Noy and Mark A. Musen. The PROMPT Suite: Interactive Tools for Ontology Merging and Mapping. *International Journal of Human-Computer Studies*, 59(6):983–1024, 2003.

[3] Sangun Park and Wooju Kim. Ontology Mapping between Heterogeneous Product Taxonomies in an Electronic Commerce Environment. *International Journal of Electronic Commerce*, 12(2):69–87, 2007.