An Automated Approach for Product Taxonomy Mapping in E-commerce

Damir Vandic, Erasmus University Rotterdam - Econometric Institute, Willem Barendszoonstraat 3, 3134 LT, Vlaardingen, Netherlands, vandic@ese.eur.nl, Flavius Frasincar

In recent years, the Web has increased dramatically in both size and range, playing an increasingly important role in our society and world economy. To address this problem, the Semantic Web was conceived to make the Web more useful and understandable for both humans and computers. One of the goals of the Semantic Web is aggregating information across multiple heterogeneous Web sources. A common aggregation problem in e-commerce is the mapping of one product taxonomy to another.

We introduce the 'Semantic Category Hierarchy for E-commerce Mapping Algorithm' (SCHEMA) to be used for mapping between heterogeneous product taxonomies from multiple sources. The main objective for developing SCHEMA is to facilitate the aggregation of product information from different Web sources and thus reducing search failures when users shop online. In order to achieve this, SCHEMA employs word sense disambiguation techniques, using WordNet, to find synonyms of the correct sense for the category name. Furthermore, it uses lexical similarity measures, such as the Levenshtein distance, together with structural information, to determine the best candidate category to map to. In order to evaluate SCHEMA, its performance is compared on recall and precision with the PROMPT and Park & Kim algorithms.

The performance of our algorithm was tested on three real-life datasets and compared with the performance of the PROMPT and Park & Kim algorithms. Our evaluation demonstrates that SCHEMA achieves a considerably higher average recall than the other algorithms, with a relatively small loss of precision. The average F1-score was 55.10% for SCHEMA, against 20.75% for PROMPT, and 32.52% for the algorithm of Park & Kim