

Buyer-Seller Matching: Global and Local Similarity Measures

Lu Yang*, Virendra C. Bhavsar*, Harold Boley**, and Biplab Sarker*

{lu.yang, bhavsar, sarker}@unb.ca harold.boleystnrc-cnrc.gc.ca

*Faculty of Computer Science, University of New Brunswick

**Institute for Information Technology - e-Business, National Research Council of Canada
Fredericton, New Brunswick, Canada

We have proposed a tree similarity algorithm for matching arc-labeled and arc-weighted buyer-seller (product) paronomy trees in e-Marketplaces. Although our tree representation removed the disadvantages of flat product/service descriptions (e.g., key words/phrases), we only syntactically derived binary similarity values (i.e., 0.0 and 1.0) between pairs of node labels. Therefore, we propose global and local similarity measures which produce more reasonable continuous similarity values between node labels to semantically enhance our tree similarity algorithm. First, our global (structural) similarity measure over paronomy trees is enriched by taxonomic semantics: Inner nodes can be labeled by classes whose partial subsumption order is represented as a background taxonomy tree that is used for class similarity computation. In particular, the class similarity can be defined via the weighted length of the shortest path connecting them. Second, leaf nodes can be typed and each type be associated with a local, special-purpose similarity measure realizing the semantics when computing the similarity of any two of its instances. We illustrate local similarity measures with e-Business types such as "Currency", "Date", and "Price". Price is the omnipresent factor that determines buyers' and sellers' decision-making in e-Marketplaces. We propose a range similarity measure that is justified by price-range overlaps between buyers and sellers in all reasonable cases. Previous price comparison systems only list the products/services whose prices fall into users' search ranges. However, our price-range similarity measure provides more fine-grained similarity values representing the percentage of the overlap between buyers' and sellers' preferred, maximum, and minimum prices.