ISSUES OF TOPICAL SEARCH AND ITEMS NORMALIZATION ON THE ONLINE MARKETPLACES

Yanholenko O.V., Sokolov D.V., Ivashchenko O.V.
National Technical University “Kharkiv Polytechnic Institute”, Kharkiv

The share of online retail on the web is growing constantly. E-commerce sales reached 2842 billion U.S. dollars worldwide in 2018 and the trend of their increase is quite stable. In order to support this tendency, modern online retail platforms work on various improvements like image and voice search, animations, social shopping to attract more potential users. With all their advantages and overall high usability, the top online marketplaces (such as Amazon, eBay, Walmart Marketplace, etc.) still make their clients face particular difficulties while performing the search of the desired products.

When a buyer tries to search for the product in some category and doesn’t know much about this type of goods (popular manufacturers, up-to-date characteristics), the task of selection of the appropriate item in the suggestion list of the online marketplace becomes not so easy [1]. In this case the search can be identified as the topical search meaning that a buyer doesn’t know up front his/her exact needs that have to be satisfied by the product that is supposed to be purchased. Therefore the search terms are modified during the search process while the client narrows the preferences. The issues of the search on the online marketplace are caused by the fact that e-commerce involves millions of sellers that can offer the same products. Moreover, the same seller can offer exactly the same product several times creating multiple items on the platform. Thus, the set of similar items (that correspond to the same product) is big enough to make a human search on it quite sophisticated.

The information technology (IT) of search on the online marketplaces suggested in the given work covers the following stages: items search; items normalization; items matching; assessment of items similarity; items clustering; and items ranking for the client. The previous research in this area has shown that similarity analysis and items clustering give poor results without the preliminary stage of items normalization. Therefore this work emphasizes the importance of normalization that is considered as a process of generating a uniform representation of the item with all its fields brought to the generalized form. This is achieved by applying the principles of intelligence theory to modeling the process of human comparison of items and creating a generalized pattern with all fields filled based on the data from multiple items descriptions. The normalized item representation allows further to solve the matching problem and build the clusters of similar items.

References:

1. Mauge K. Structuring e-commerce inventory. / K. Mauge, K. Rohanimanesh, J.-D. Ruvini // Proc. of the 50th Annual Meeting of the Association for Computational Linguistics. – 2012. – P. 805-814.

2. Dong Y. Normalization of duplicate records from multiple sources / Y. Dong, E.C. Dragut, W. Meng // IEEE Transactions on Knowledge and Data Engineering. – 2018. – Vol. 31. – P. 769-782.