

FORMAL METHOD FOR QUALITY DETERMINATION OF POSTS ON SPECIALIZED SITES

Oleh Bisikalo¹, Hanna Kyrylenko²

Institute for Automatics, Electronics and Computer Control Systems, Vinnytsia National Technical University, Khmelnytskeshose St., 95, Vinnytsia, 21021, UKRAINE, E-mail: ¹obisikalo@gmail.com,
²anyakurull@rambler.ru

The article is devoted to the problem of the Internet searchers of useful information. The question is very relevant today since the Internet technologies are developing rapidly, and the amount of information in the network is increasing at a fast pace. The search of necessary information takes much time. The method proposed in this article will reduce the time wasted on the search of useful posts on specialized forums.

The method of the formal post quality determination based on the array of parameters is considered in the article. The authors have proposed to use the following parameters: the post size, the links presence, the question mark presence, the ellipsis presence, and the code percentage in the post. During the research it has been defined that all the proposed parameters are informative for the post usefulness determination. The characteristics are received from the parsing of the HTML-code of the page with posts via Java library Jsoup. The next stage is the decision tree building using the Matlab tools. At the input the decision tree receives the text file with the parameters array for each post. At the output we get the answers array whether the post is useful or not.

In the authors' opinion the advantage of the method of is its simplicity. Since the HTML-code is the input text, there is no need to mark it out additionally. And we use the prepared library Jsoup for parsing. Giving the high percentage of correct post quality indications, the proposed method will allow users to reduce the time spent on the necessary information search.

Keywords – post, parsing, Jsoup, Matlab, decision tree.