Contribution ID: **324** Type: **not specified**

Generating record templates as a subtask for extracting entities from poorly structured data, using author affiliation information as an example

Thursday, 6 July 2023 17:30 (15 minutes)

The study is devoted to developing an algorithm for extracting the names of organizations from poorly structured data. Bibliographic information about the publications from the abstract database Scopus was taken as the initial data.
The main problem in extracting names of organizations from affiliations, apart from the presence of typos, is that the requirements of journals and conferences to spell affiliations are different. This results in affiliations to the same organization being written in different ways, which does not allow for statistical analysis on organizations. In this regard, the authors of the research analyzed 750 records with affiliations of the publication's authors and used them for statistical analysis of affiliation writing templates and compiled a list of the 10 most frequently used ones (186 different templates in total). Based on the templates compiled, an algorithm was developed to identify the names of organizations.
In order to analyze the effectiveness of this method, the authors of the study conducted an experiment comparing the accuracy of identification of the name of the organization using two algorithms: one developed without templates and one developed on the basis of templates. The results of the experiment confirm the effectiveness of the template method for further development of the algorithm before developing it without the use of templates.

Summary

Primary authors: FILKIN, Ivan (National Research Nuclear University MEPhI); ULIZKO, Mikhail (National Research Nuclear University MEPhI); TUKUMBETOVA, Rufina (Plekhanov Russian University of Economics)

Presenter: FILKIN, Ivan (National Research Nuclear University MEPhI)

Session Classification: Big Data, Machine Learning and Artificial Intelligence

Track Classification: Big Data, Machine Learning and Artificial Intelligence