

Text Classification Task in Frames of Big Data Processing

M.Khachidze, M. Tsintsadze, M.Archuadze, G.Besiashvili

e-mails: {[manana.khachidze](mailto:manana.khachidze@tsu.ge), [magda.tsintsadze](mailto:magda.tsintsadze@tsu.ge),
[maia.archuadze](mailto:maia.archuadze@tsu.ge), [gela.besiashvili](mailto:gela.besiashvili@tsu.ge)}@tsu.ge,

Computer Science Department, Faculty of Exact and
natural Sciences , I.Javaxishvili Tbilisi State University.
Chavchavadze Ave. 3 Tbilisi, Georgia

In a universally recognized big data classification task, the data generated from social networks is in the fifth place. This fact increases the importance of the natural language processing (NLP - Natural language Processing) in frames of information retrieval from large volumes of unstructured, semi-structured and structured data. From several tasks of NLP the problem of information retrieval is quite important, especially in case of multilingual data processing. The studies on this direction are impressive for “multiuser” languages, but on the other hand are not sufficient for "small" (minor) languages.

We consider the so-called free-text classification task as a subtask of information retrieval for Georgian Language based texts. In the work provided we describe the method allowing classification of texts of various type, volume or hierarchy, irrespective of their sources being it structured database, language building, text collection or social networks.

We will also describe the possible applications of this method for medical records, as a data processing tool for biomedical research in bioinformatics.

References

- [1] J.Monti, M. Monteleone, M.P. Buono, F. Marano, "Natural Language Processing and Big Data - An Ontology-Based Approach for Cross-Lingual Information Retrieval", vol. 00, no. , pp. 725-731, 2013, doi:10.1109/SocialCom.2013.108 .
<http://doi.ieeecomputersociety.org/10.1109/SocialCom.2013.108>
- [2] J.Sun, C. K. Reddy. Big Data Analytics for Healthcare. Tutorial presentation at the SIAM International Conference on Data Mining, Austin, TX, 2013.
<https://www.siam.org/meetings/sdm13/sun.pdf>
- [3] E.Kha, Addressing Bioinformatics Big Data Problems using Natural Language Processing: Help Advancing Scientific Discovery and Biomedical Research. Modern Computer Applications in Science and Education. p.221-228. <http://www.wseas.us/e-library/conferences/2014/CambridgeUSA/COMP/COMP-31.pdf>
- [4] M.khachidze, M.Tsintsadze, M. Archuadze. Natural Language Processing (NLP) Based Instrument for Classification of Free Text Medical Records. BioMed Research International Volume 2016 (2016), Article ID 8313454, 10 pages.
- [5] <http://dx.doi.org/10.1155/2016/8313454>