

## ტექსტების კლასიფიკაციის ამოცანა დიდი მონაცემების დამუშავების ჭრილში

*მ.ხაჩიძე, მ.ცინცაძე, მარჩუაძე, გ.ბესიაშვილი*

ელ-ფოსტა: {[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,

კომპიუტერულ მეცნიერებათა დეპარტამენტი,  
ზუსტ და საბუნებისმეტყველო მეცნიერებათა  
ფაკულტეტი, ივ.ჯავახიშვილის სახელობის  
თბილისის სახელმწიფო უნივერსიტეტი  
თბილისი. ჭავჭავაძის გამზ. 3

საყოველთაოდ აღიარებულ დიდი მონაცემების კლასიფიკაციაში სოციალური ქსელებით გავრცელებულ მონაცემებს სიდიდით მეხუთე ადგილი უკავია. ეს ფაქტი ზრდის ბუნებრივი ენის დამუშავების (NLP – Natural language Processing) მნიშვნელობას ინფორმაციის ამოღების უზრუნველსაყოფად დიდი მოცულობის არასტრუქტურირებული, სტრუქტურირებული და ნაწილობრივ სტრუქტურირებული მონაცემებიდან. NLP-ს სხვადასხვა ამოცანებს შორის ინფორმაციულ ძებნა ერთერთი მნიშვნელოვანია, განსაკუთრებით მრავალენოვანი მონაცემების არსებობისას. ამ მიმართულებით კვლევები ერთის მხრივ მრავლადაა „დიდი“ (მრავალმომხმარებლიანი) ენებისათვის, მეორესმხრივ განიცდის მწვავე უკმარისობას „პატარა“ (მცირემომხმარებლიანი) ენებისათვის.

ჩვენ განვიხილავთ ეგრეთწოდებული თავისუფალი ტექსტების კლასიფიკაციის ამოცანას, როგორც ინფორმაციული ძებნის დიდი ამოცანის ქვეამოცანას ქართულენოვანი ტექსტებისათვის. აღწერთ მეთოდს, რომლის საშუალებითაც შესაძლებელია სხვადასხვა ტიპის, შინაარსობრივი იერარქიის და მოცულობის ტექსტების კლასიფიკაცია განურცევლად იმისა, თუ რა იქნება ამ ტექსტების წყარო - სტრუქტურირებული მონაცემთა ბაზა, ენის კორპუსი, ტექსტების კოლექცია თუ სოციალური ქსელები.

აღწერთ შესაძლებლობას ამ მეთოდის გამოყენებისა სამედიცინო ჩანაწერებზე, როგორც ბიოინფორმატიკაში დიდი მონაცემების დამუშავების ინსტრუმენტი ბიოსამედიცინო კვლევების მხარდასაჭერად.

### ლიტერატურა

- [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>
- .