

In this study we seek to reveal topic structure of ethnicity related discussion in the Russian language social media, to compare public attention to different ethnic groups, and to develop an approach that would automatically detect various aspects of attitudes to those ethnic groups.

115 ethnic groups

“indigenous” to the post-Soviet space were investigated.

4063 ethnonyms

were generated for this groups (bigrams and pejorative ethnonyms are included).

>2.6 million of texts

containing at least one keyword from more than 10,000 Russian-language social media were gathered for a two-year period from 2014 to 2015.

15000 texts manually labeled

using a set of 15 attitude-related questions and three coders per each text.

Based on the hand-coded sample we find that the negative end of attitudes, apart from being dominated by various ethnofaulisms, diverges from the data obtained in public opinion polls. Not Caucasians, but various Central Asians take the lead in negativity. As for Caucasians, it is they who most often write for themselves – that is, produce their own discourse which is most likely to shift their scores up. It is also not a surprise that Ukrainians are among most negatively represented because of the recent military conflict. These conclusions generally match with our earlier research on a sample of popular bloggers. We conclude that to expand this research we need to improve the instrument of automatic attitude detection by shifting to the sentence-level analysis.

97 topics

were defined to characterize the global semantics of the ethnical discussion (LDA). We find that the most salient topics are the five ones concerning Ukrainian-Russian relations and military conflict at the Ukrainian’s East. We also find that some ethnicities are significantly more associated with certain topics – in particular, “Ingush”, “Mansi”, “Caucasian” are significantly more probable in the crimes and murders topic.

Automatic detection of attitudes

For feature selection, a lexicon of bigrams was trained on the main sample of 2.6 million texts with Gensim Phrases module. As the target variables were the mean assessors’ scores. Then we train SGDClassifier with scikit-learn. In brackets – gain over classifier that randomly assigns classes to texts / instances, albeit keeping the true class proportion.

Text-level prediction

General negative sentiment
General positive sentiment
Interethnic conflict
Positive interethnic interaction

Instance-level prediction

Attitude (pos, neg, neutral)
Victim / aggressor
Inferior / superior

Precision

0.75 (+14)
0.72 (37)
0.71 (+24)
0.70 (+35)

Recall

0.73 (+21)
0.64 (+6)
0.71 (+19)
0.62 (+5)

F₁

0.74 (+17)
0.65 (+17)
0.71 (+21)
0.63 (+15)

0.67
0.62
0.67

0.55
0.61
0.67

0.58
0.61
0.67

Acknowledgements: This work was done at the Laboratory for Internet Studies, National Research University Higher School of Economics (NRU HSE), Russia, and supported by the Russian Research Foundation grant no. 15-18-00091.