The Migration Topic in the Russian-Language Sector of LiveJournal

Svetlana Popova
Saint-Petersburg State University, Saint-Petersburg, Russia
ITMO University, Saint-Petersburg, Russia
svp@list.ru

Vera Danilova
Russian Presidential Academy of National Economy
and Public Administration
maolve@gmail.com

Abstract—**This paper contains the results of the analysis of several thousand LiveJournal (LJ) posts on migration. The data for the analysis has been gathered using search queries.**

I. Introduction

Social networks experienced a genuinely vast development. They became one of the main sources of public opinion and an important tool for grading its formation. Our study aims, on one hand, to consider the representation of migration topic in a collection of posts from a given social network, and, on the other hand, to detect outside influences on the public opinion, if any. This paper reports on a pilot study of LJ search results that include several thousand entries on migration (without comments). The study includes the following stages: 1) data collection, 2) pre-processing, 3) a manual analysis of posts with high and low (from 3 to 10 reposts) repost level, and analysis of topic (phrases) frequency characteristics, 4) performance evaluation. The findings of this research are thoroughly described in [1]. The language for queries and posts was Russian.

II. Input dataset

We have selected LJ search results as the source for data collection. The search output represents answers to a number of queries that have been worded and grouped. The queries have been picked up manually by analyzing the search output for each of the queries and leaving only those containing less noise (posts unrelated to the considered topic). Table 1 shows conditional names for query groups and the queries for each of these groups, as well as the number of posts per group with and without the account of reposts.

III. Keyphrase construction

In the paper we extract topics as phrases. One of the main approaches to keyphrase extraction includes two stages: (i) construction of a candidate list, and (ii) candidate ranking and selection of a predefined number of best phrases or classification of candidates as being either keyphrase or not a keyphrase. Instead of ranking/classification we assume to employ an extended stop words list that allows to remove most general-use expressions. The use of such lists raise the quality of extracted phrases [6-8]. Also it has been shown in [2-5], part-of-speech tagging and, particularly, adjective and noun annotation, plays a significant role in solving the keyphrase extraction problem.

TABLE I. Queries and query groups

Query group	Queries entering this group	The number of posts	The number of posts without duplicates
State and politics	illegal immigrants+migration, migration+problems, immigrants+problems	1183	567
Migration and residence	residence+of+immigrants population+immigrants, population+migration	929	536
Migration and work	economics+migration, economics+immigrants, migration+politics	1115	481
Cultural adaptation, Tolerance	migration+tolerance, migration+ethnicity	463	217
Migration and crime	migration+ethnic+crime, immigrants+crime	454	94

In this work, we use keyphrases consisting of sequences of maximum length of nouns and adjectives and an extended stop word list to perform automatic topic (keyphrase) extraction. Stop words are delimiters between phrases in this case, also as punctuation marks and words of other parts of speech. Together with standard stop words like on (he), v (in), i (and), na (on), etc. our list includes terms that occurred in the list of candidate phrases with high frequency, however, which are not context-specific (e.g., to the extent that, soon, confirmation, option, a few, etc.). These terms were selected manually from a list of frequent phrases constructed previously using a standard stop words list. The inclusion of the above frequent context-independent words to the extended stop words list allowed for an improvement in the quality of the phrases (cheked manually).

IV. Main obtained results

Most messages from the 2014 collection were posted in the months of February, May and September. We consider the posts having a considerable number of shares (over 10), as well as post groups shared between 3 and 10 times.Additional the main topics of all posts have been explored. Reposts are not taken into account during topic extraction.

The manual examination of posts having most shares indicates that:

1) a considerable part of these posts is related, directly or indirectly, to the Moscow elections;

2) most part of these messages contains criticism of the roulette of illegal migration (commonly directed to rotten bloggers, mass media, opposition candidates);
3) the majority of reposts emphasize the positive sides of the migration: victories of immigrants in competitions, the Russian economys need of immigrants, initiatives to establish closer relations between immigrants and population, hard-working nature of immigrants, the relation of the overall crime level to the count of criminal acts committed by immigrants, criminal acts of compatriots toward immigrants;
4) these posts often draw attention towards the actions of the authorities intended to calm the situation, as well as to the efficiency of the Federal migration service in the previous period;
5) very few messages show a strong negative reaction towards immigrants, talking about the substitution of Russians by immigrants and negative sides of migration (considering immigrant from the Central Asia)

Our impression from the study of the highly shared posts is that there are two main accents made and both are pro-government:

1) smoothing anti-immigration attitudes: immigrants are necessary for the economic stability of the country and they really can bring a positive contribution (for example, winning the competitions), successful initiatives and pronouncements of the authorities (and/or the Federal migration service) are emphasized, related to the migration issues in a way that should be received positively by the citizens;
2) appeals not to respond to the calls of the opposition that uses one of the most sensitive issues for the Moscow citizens in the election race.

It can be assumed that both accents are due to the pre-election propaganda, despite the fact that in some messages it is not explicitly mentioned, and the posts themselves appear to be published during a calm period. The event of reposting certain messages that might seem unremarkable at the first glance sometimes cannot be easily explained and we suppose these shares are propaganda-oriented. It stands to mention that reposts practically do not mention complex topics, such as the issue of visa regime with the countries of Central Asia, which allows us to suppose that the above mentioned reposts are made on purpose and pursue the goals of the election campaign. In this case, one can judge indirectly about the issues sensitive for the population. It is the mitigation of these problems that is usually addressed by the posts that are intended to defuse the negative attitudes towards immigrants.

The titles of the entries reposted less than 10 times cover a rather wide range of topics:

1) migration in Greece and other EU countries;
2) migration in Belarus;
3) migration in the US;
4) conference debates on migration, etc.

The statistics of all posts (excluding reposts) shows that inside the gathered collection the most popular are the topics (phrases) related to labour migration, migration policies, migration service, visa regime, adaptation of immigrants and the Russian language. The discussions are primarily concerned with the immigrants from the Central Asia. According to the obtained statistical descriptions, the cultural adaptation issues are less prominent in the considered posts than those of labour migration and illegal immigrants.

It stands to mention that, as the results reported here are based on the LJ search output and correspond only to the information seen by the user in response to his queries and other topically close queries, we can only make indirect conclusions on the structure and contents of a discussion.

It is interesting to note the following contrast. In [9-11], another approach to data collection has been employed based on a regular gathering of posts belonging to the top 2000 bloggers, according to the LJ ranking, where the position of a blogger is calculated as a function of the number of his/her friends taking into account the reading activity of these friends [11]. At the next stage, topic-specific posts are sorted out using one of the topic modeling algorithms [9-11, for details see 11]. In [9], the election campaign of 2011-2012 is considered. On the basis of the results the following conclusion is made: an important empirical finding of this study is that there is a quantitative prove of the fact that the most influential Russian blogs play the role of a mass media stronghold of the opposition. A 13-weeks long analysis of posts shows an absolute political dominance of opposition bloggers [9].

In [12], the migration topic arises when considering a cluster that contains a number of Russian pro-nationalist bloggers: from extremists that propagate violence towards immigrants from the Caucasus and Central Asia to less radical nationalists that are focused on the Russian and Soviet history, Russian orthodox church and football The movement against the illegal immigration is the only socio-political movement related to this cluster. Here we would like to remind that the posts with multiple shares in 2014 do not criticize authorities and immigration almost at all, on the contrary, they benefit the upcoming election campaign and highlight the positive aspects of migration, avoiding critical issues. This contrast can be considered an indirect evidence of the strategies of the governing authorities and opposition to promote their position, or of the increasing activity of pro-government bloggers.

REFERENCES

[1] Popova S., Egorov A., Khodyrev I., Danilova V., Topic analysis of LiveJournal posts on the theme ”migration”. / mathematical modeling and informatics of social processes. 2015. P 156-178 (in Russian)

[2] Hulth A. Improved automatic keyword extraction given more linguistic knowledge. In: Proc. of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP ’03). 2003. P. 216223.

[3] Mihalcea R., and Tarau P. TextRank: Bringing order into texts. In: Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP ’04). 2004. P. 404411.

[4] Wan X., and Xiao J. Exploiting Neighborhood Knowledge for Single Document Summarization and Keyphrase Extraction. In: ACM Transactions on Information Systems. 2010. V. 28. N 2. Article 8.

[5] Zesch T., and Gurevych I. Approximate Matching for Evaluating Keyphrase Extraction. In: Proc. of the International Conference on Recent Advances in Natural Language Processing (RANLP 2009). 2009. P. 484489.

[6] Popova, S., and Khodyrev, I. Ranking in keyphrase extraction: is it useful to take into account the frequency characteristics of candidate phrases? In: Proceedings of the Institute for System Programming of RAS, vol. 26, No. 4, 2014

[7] Popova, S., and Khodyrev, I. Using keyphrase extraction and ranking for annotation purposes. In: Scientific and technical journal of information technologies, Mechanics and Optics, No. 1 (83), P. 81-85, 2013

[8] Popova S., Kovriguina L., Muromtsev D., and Khodyrev I. Stop-words in Keyphrase Extraction Problem. In: Proc. of 14th Conference of Open Innovations Association FRUCT. Helsinki, Finland, 2013. P. 113121.

[9] Koltsova, O., and Shcherbak, A. LiveJournal Libra! The influence of the political blogosphere on political mobilisation in Russia in 2011-12 // Paper submitted to New Media and Society

[10] Mitrofanova, O., Shimorina, A., Koltsov, S., and Koltsova, E. Modeling semantic links in social media texts with LDA (a case study of the Russian-language LiveJournal). In: Structural and applied linguistics, No. 11, St. Petersburg: 2014 (in Russian)

[11] Kotsova, O., and Koltsov, S. Statistical and topic profiles of LiveJournal. In: Proceedings of the XVI all-Russia Scientific Conference Internet and Modern Society (IMS-2013), St. Petersburg, 2013 (in Russian)

[12] Etling, B., Alexanyan, K., Kelly, J., Farid, R., Palfrey, J., and Gasser, U. Public discourse in the Russian blogosphere: mapping RuNet politics and mobilization. In: Studies of the Berkman Center, No. 2010-11, 2010