Objectified Formulation of a Search Query for an Intellectualized System of Monitoring Internet Memes*

Oleg A. Gabrielyan [1][0000-0003-0302-0229], Mikhail V. Gasparyan [2][0000-0002-8183-8475], Ivan V. Kravchenko [3][0000-0001-6661-5698]

[1] V.I. Vernadsky Crimean Federal University, Prospekt Vernadskogo, 4, Simferopol, Crimea, 295007, Russia,
gabroleg@mail.ru

[2] .I. Vernadsky Crimean Federal University, Prospekt Vernadskogo, 4, Simferopol, Crimea, 295007, Russia,
mgasp@mail.ru

[3] V.I. Vernadsky Crimean Federal University, Prospekt Vernadskogo, 4, Simferopol, Crimea, 295007, Russia,
zx3com@gmail.com

Abstract. The evolution of information technologies has currently reached the point allowing us to state the fact of formation of a new type of society, extending further on from informational to digital. The pivotal role in its formation belongs to virtual communications based on technical and software tools created by man. Novel channels of virtual communication greatly surpass the natural forms of communication in terms of the amount of transmitted information, which significantly affect the processes of socialization and orientation of a person under present conditions. Among important phenomena embodying the features of informational, virtual processes characteristic for the 21st century, are Internet memes, representing information units in the modern world. Internet memes in a virtual space accumulate a significant potential for influencing public opinion. In the paper, we overview the procedure for generating a query for a Google search engine in the context of designing an intellectualized monitoring system for Internet memes. The procedure of a search query generation is implemented based on the analysis of the two information sources: the first one containing data on relevant sociopolitical agenda and the second one – on the public perception and interpretation of this information. The proposed procedure of query generation has certain advantages in that it can easily be translated into a relevant software code and takes into account the dynamics of social and political processes in Russia.

Keywords: Internet Meme, Search Query, Social and Political Processes, Database of Political Internet Meme Images.

1 Introduction

Modern-day communications have largely moved to a virtual environment. The latter still awaits its proper examination because such communications and their content greatly affect undergoing social processes.

Internet memes belong to a form of such communications. The overall quantity of Internet memes, which can be defined as semi-structured datasets, keeps rising steadily [1].

What we deal herewith is an open and constantly replenishing system, not only in terms of its content, but also in terms of genre diversity (slogans and Photoshop-memes, viral videos and parodies, melodies and gifs, comic strips and demotivators), and semiotic characteristics (verbal (text), auditory, visual and mixed messages) [2].

Internet memes in the virtual place possess a big potential to influence public opinion. Hence the importance of producing algorithms (to be then implemented in software products) designed to monitor the processes that determine their origin, structure, and distribution in an automated mode. Among other things, this is because processing the data stream of Internet memes, that is, the big data analytics can be hardly conducted by people on their own. Too much information is needed to get processed and comprehended [3].

Possessing a viral nature of distribution, Internet memes not only "infect" communicants, but also change their thoughts, attitudes, and behavior [4]. Social practices show that the most efficient way to manipulate people's minds is to influence their deep meanings – the archetypes that make up the content of the collective unconscious. This idea is very well understood and effectively used in marketing, politics, and ideology [5]. Concerning Internet memes, this is mostly related both to their presentation forms (striking and memorable "packaging") and emotional effects which address rather one's sensual than logical perception. That said, memes are targeted to the collective unconscious (archetypes, cultural codes) and the culture of laughter. The dynamics of "viral processes" in society has already received its descriptions and mathematical models [6, 7], and we see it now necessary to implement them in creating an automated system for the analysis of influence (monitoring) rendered by Internet memes in a virtual environment on the Russian-speaking Internet segment [8, 9, 10].

In the paper, we tackle the procedure for generating a query, taken as a component of the intellectualized system for monitoring Internet memes.

Supported by the Russian Foundation for Basic Research (№ 20-011-31460), the project "Political meme in virtual communications: software product design for monitoring its distribution and influence (analysis of the Russian-speaking Internet segment)" has also the aim, among other things, to make a primary database of political Internet memes.

The paper's goal is to describe the procedure of formulating a search query necessary for creating the database for political Internet meme images.

When formulating the search query aimed at building up the primary database named "The database of political Internet meme images" (IDB), it was principal for us to design such a mechanism that could be easily translated into computer software (SW).

Our task was to find a constantly and automatically updated source of information reflecting relevant social and political processes in Russia.

2 Search Query Formulation Procedure

When formulating the search query, we used the two information sources complying with the noted requirements. The first source of information included the headlines for the news of sociopolitical nature over a certain period taken from the newspaper Kommersant (Businessman). In the fully functional version, when formulating queries, feeds from news aggregators will be analyzed. A set of news headlines collected over some time forms the hypertext.

The second source contained the data from the weekly opinion poll "FOMnibus" (Public Opinion Foundation). In particular, it was analyzed an array of data extracted from respondents' answers to an open question: "What events of the past week, which were reported in mass media, seemed most interesting to you or attracted your attention?" [11]. The use of this array helped counterbalance the language by the official media by complementing the query with informal slang (everyday language). This was important since the object of our study is partly a folklore product – Internet folklore. Besides, it consists of an array of tags that can be immediately used in the description and the search for political memes, in that helping to build the primary database – IDB (image database).

The query itself consists of two parts. The first one is to reflect and define the subject of a judgment (query predicate). Within this work's framework, the predicate will be defined as a constant term, namely, meme. The second part is variable and reflects the studied subject in the structure of virtual communication. To receive search results closest to the content of the query, the query wording is placed in quotation marks ("meme …") [12].

To fulfill the study, the Google search engine will be used, as, according to many objective indicators, it addresses the most complete database of images, while the algorithm of query making appears to be easy to get formalized.

3 Analysis of Empirical Data

To test the proposed approach, let us set the period as follows: from March 01 to March 31, 2020.

Based on the news feed archive from the Kommersant newspaper [13], a hypertext is produced for the indicated period. Next, the resulting hypertext is subjected to the onto-semantic analysis using the software application KH Coder [14]. Relevant results are shown in the form of a semantic network [15] in Figure 1 (here and throughout the figure the terms are presented in the original language. Further in tables 1 to 4, is given the translation of the concepts of vertices of the graphs analyzed in this work).

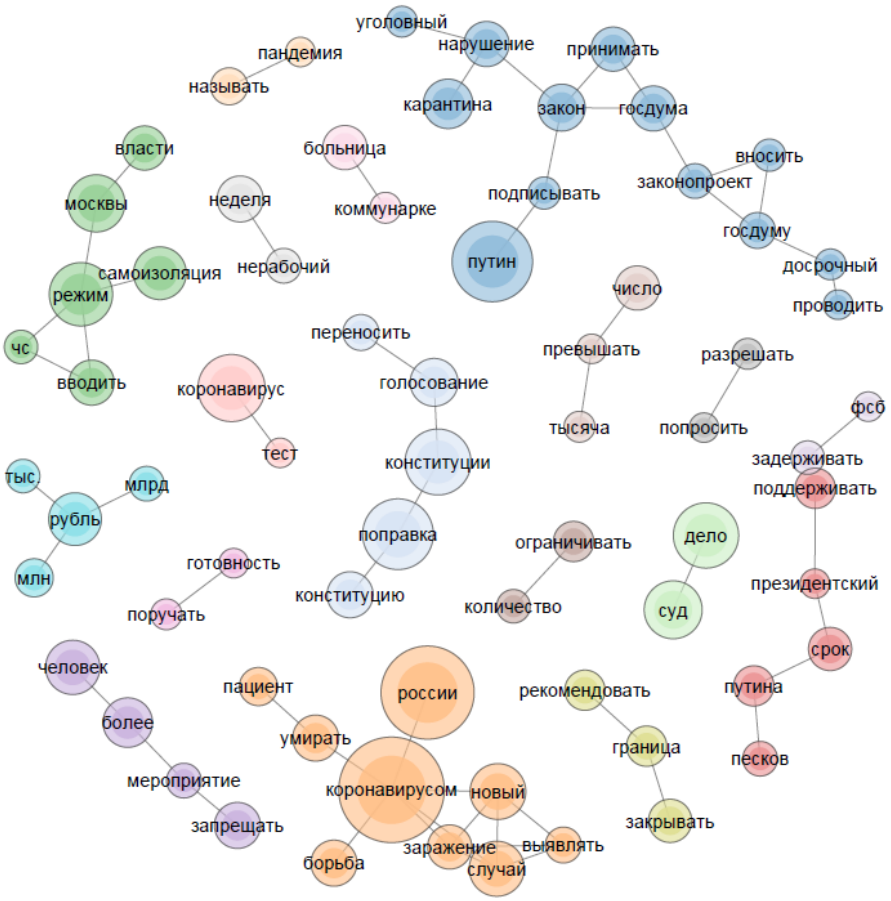

Fig. 1. The semantic network built for the hypertext relevant to the period from March 01 to March 31, 2020.

As a next step, let us go up to the analysis of the constructed graphs, the aggregate of which defines the semantic network of the hypertext for the given period. To describe the principle of word selection when formulating a query, let us consider two graphs, whose choice is arbitrary. Figures 2 and 3 show the examples of graphs built for the semantic network of the hypertext. Let us denote them according to the names of the corresponding graph vertices: “Coronavirus” (graph in Figure 2) and “Regime” (graph in Figure 3).

Fig. 2. Graph «Coronavirus».

Fig. 3. Graph «Regime».

The query is composed of words associated with the vertex (word) having the highest degree in the graph. To do so, we build a ranked sequence of degrees for the graph's vertices (Table 1, Table 2). The query includes only the words (vertices) adjacent to the vertex having the greatest degree.

Table 1. Vertices of the graph "Coronavirus" ranked according to their degree.

Vertex	Degree
Coronavirus	5
New	4
Cases	4
Infection	3
Russia	1
Die	1
Fight	1

Table 2. Vertices of the graph “Regime” ranked according to their degree.

Vertex	Degree
Regime	4
Emergency	2
Introduce	2
Moscow	1
Self-isolation	1

Based on the data, shown in Table 1, we can formulate the following queries: “Coronavirus new case infection Russia”; “Coronavirus new case infection die”; “Coronavirus new case infection fight”.

Based on the data, shown in Table 2, we can formulate the following queries: “Regime introduce emergency self-isolation”; “Regime emergency introduce self-isolation”; “Regime introduce emergency Moscow”; “Regime emergency introduce Moscow”.

Table 3 shows the query wordings and image search frequency results when the Google search engine was used. Found images can be downloaded by using the Download All Images tool. Next, the collected images were examined to eliminate duplicates with the help of the Find.Same.Images.OK application. This application seeks duplicate images by comparing them at the pixel level. This allows for efficient comparative analysis and filtering of the same images, even after they have been edited (rotated, mirrored, cropped, etc.). As a result, based on 1119 images, we accumulated 525 unique sets of queries.

Table 3. Frequency distribution for the search query results (source 1).

Query	Search results (images, items)
Regime introduce emergency self-isolation	184
Regime emergency introduce self-isolation	176
Regime introduce emergency Moscow	173
Regime emergency introduce Moscow	176
Coronavirus new case infection Russia	180
Coronavirus new case infection die	192
Coronavirus new case infection fight	118
Total	**1119**

For the whole four weeks in March 2020, the first place in the ranking of the most important events, according to the opinion poll "FOMnibus", was occupied by the Coronavirus epidemic, Coronavirus pandemic. Based on all the respondents’ answers, specific semantic units were selected, then formulated in the form of search queries. Table 4 shows a set of queries and frequencies of their use when searched in the Google search engine. The primary collection of these queries consisted of 2,699 images. On

fulfilling the procedure of comparison and elimination of duplicates, 1401 images remained.

Table 4. Frequency distribution for the search query results (source 2).

Query	Search results (images, items)
coronavirus monitored	188
virus from China	116
coronavirus spreading Europe	173
Coronavirus tired of artificially made	182
epidemic of Coronavirus	141
Coronavirus, pandemic	181
disease terrible, Coronavirus	171
Coronavirus, phobia disease	170
the threat of Coronavirus infection	182
Coronavirus, widespread in Russia	183
Coronavirus: soon quarantines everywhere	167
virus – panic	167
closure of borders	168
school quarantines	184
psychosis pumped up due to Coronavirus	156
Coronavirus, artificially created panic	170
Total	**2699**

Table 5. Comparison of the search results for the two sources.

	Search results (images, items)	Result after comparison (images, items)	Share of unique images
Source 1	1119	525	47%
Source 2	2699	1401	52%

Judging by the data in Table 5, extending the structure of search queries (based on the Source 1, we generated 7 queries and based on the Source 2 – 16 queries) resulted, on the one hand, with a quite obvious consequence – an increase in the number of search results (images), and, on the other hand, an increase in the number of the unique search results. To sum up, the structure of queries makes an impact on the statistics of search results.

4 Conclusion

Having aggregated the arrays of images collected through the use of the queries formulated with the help of two sources, we obtained 1926 images. The procedure of their

comparison reduced their number down to 1791 items. As a result, we created an IDB ready for further analysis and defined the mechanism of formulating queries.

This database requires to be first filtered according to certain objective criteria, which must be specified and formally described. This is necessary to automatize the selection process for a more relevant IDB. The proposed method of formulating queries for its further processing has an advantage in that it is done objectively and can be translated into an automatized software application without the direct intervention of a researcher or an expert.

The reported study was funded by RFBR and EISR according to the research project № 20-011-31460.

1 References

1. Gabrielyan, O.A., Lukyanenko, V.A., Kozlova M.G., Gasparyan, M.V., Gabrielyan, T.O.: Intellectualization Of The Sociometric Data Processing Of Internet Memes Within Virtual Communication Structure. International Scientific Forum National Interest, National Identity, and National Security. The European Proceedings of Social & Behavioural Sciences.
2. Savitskaya, T.E.: Internet Memes as a Phenomenon of Popular Culture. Retrieved from: http://infoculture.rsl.ru/donArch/home/KVM_archive/articles/2013/03/2013-03_r_kvm-s3.pdf.
3. Gabrielan, O., Suleimanov, I., Vituleva E.: Artificial Intelligence in the Context of Noosphere Studies. International Scientific Conference Social and Cultural Transformations in the Context of Modern Globalism. The European Proceedings of Social & Behavioural Sciences, pp. 975-981.
4. Rashkoff, D.: Media Virus. Hidden Agenda in Popular Culture. Moscow, Ultra. Kultura, (2003). (In Russian)
5. Gabrielian, O. A. Mythopoetics of the Culture: the Research Methodology Issue ASTRA Salvensis, Supplement no. 1, pp. 549-560 (2020)
6. Suleimenov, I., Gabrielyan, O., Bakirov, A., Vitulyova, Y.: Dialectical Understanding of Information in the Context of Artificial Intelligence Problems. Published under license by IOP Publishing Ltd/ IOP Conference Series: Materials Science and Engineering, Volume 630, Number, pp.1-8. (2019)
7. Suleimenov, I.E., Vitulyova, Y.S., Bakirov, A.S., Gabrielyan, O.A.: Artificial Intelligence - what is it? 6th International Conference on Computer and Technology Applications (ICCTA 2020), published by the Association for Computing Machinery, pp. 22-25. (2020)
8. Germanchuk, M.S., Kozlova, M.G., Lukyanenko, V.A.: The Problems of Modelling Internet Meme Distribution Processes. Analysis, modeling, management, design of socio-economic systems: proceedings of the XII International school-symposium AMUR-2018. Simferopol: IP Kornienko A.A., P. 136-139. (2018). (In Russian)
9. Germanchuk, M.S., Kozlova, M.G.: Recognition, Analysis, and Visualization of Internet Memes. Mathematical methods of image recognition: Proceedings of the 19th All-Russian conference with international participation. M.: Russian Academy of Sciences, P. 351-355. (2019). (In Russian)
10. Germanchuk, M.S., Kozlova, M.G., Lukyanenko, V.A.: Development of tools for processing and analysis of the Internet meme data stream. Proceedings of the All-Russian scientific and practical conference MIKMO-2019 and Taurida scientific school conference for

students and young specialists in mathematics and informatics / Ed. V.A. Lukyanenko. – Simferopol: IP Kornienko A.A. V 1. P. 121-127. (2019). (In Russian)
11. Public Opinion Foundation: Dominants. Opinion Field. Events of the Week. V 9-12.
12. Information – Google Search: website. URL: https://support.google.com/websearch/answer/2466433 (date of reference: 24.07.2020).
13. Kommersant: URL: https://www.kommersant.ru/archive/news/month/2020-03-01?nodesid=47 (date of reference: 24.07.2020)
14. KH Coder 3 Reference Manual. URL: https://khcoder.net/en/manual_en_v3.pdf (date of reference: 30.07.2020). [website… Text: electronic]
15. Panchenko, A. et al.: Human and Machine Judgements for Russian Semantic Relatedness. Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016, Yekaterinburg, Russia, pp. 221–235. DOI: 10.1007/978-3-319-52920-2_21.