LINGUISTIC ANALYSIS OF INTERNET MEDIA AND SOCIAL NETWORK DATA IN THE PROBLEMS OF SOCIAL TRANSFORMATION ASSESSMENT*

M. Zgurovsky,[1][†] D. Lande,[2] A. Boldak,[1][‡]
K. Yefremov,[3] and M. Perestyuk[4]

UDC 303.732+004.62+004.912+351/354

Abstract. *A combined approach has been developed to estimate the effectiveness of social transformations as a measure of inconsistency between the actions of the authorities and the expectations of society and the synergy (social activity) of people, based on formalized coordination of the results obtained by the method of expert assessments and methods of sentiment analysis and intelligent analysis of text messages from online open sources and social networks. These methods are implemented as a set of web services and applications in the development environment of the Advanced Analytics integrated online platform of the World Data Center "Geoinformatics and Sustainable Development." The effectiveness of the proposed approach is demonstrated by a quantitative assessment of the attitude of the population of Ukraine to the actions of the authorities concerned with the spread of the COVID-19 pandemic.*

Keywords: *vector of government actions, vector of society's expectations, vector of transformations (reforms), linguistic analysis, content analysis, linguistic sentiment analysis of Internet media data and social networks, open-source intelligence.*

GENERAL APPROACH TO ASSESSMENT OF SOCIAL TRANSFORMATION EFFICIENCY

The interaction of authorities and the public is very relevant for any country, its part, or any large corporate or public entity. A quantitative assessment of public support for government actions is a much needed instrument in substantiating managerial decisions including potential conflicts and critical situations.

Assume that the social transformations (reforms) being assessed by the social transformation factor R are insofar as efficient as the vector of expectations and social activity of society $\vec{S}$ comes close to the vector of government action $\vec{G}$ by the direction and how high and positive is the synergy of society (Fig. 1). By the synergy of society, we mean

[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine, [†]*zgurovsm@hotmail.com*; [‡]*boldak.andrey@gmail.com*. [2]Institute for Information Recording, National Academy of Sciences of Ukraine, Kyiv, Ukraine, *dwlande@gmail.com*. [3]World Data Center for Geoinformatics and Sustainable Development of National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine, *k.yefremov@gmail.com*. [4] Institute for Monitoring the Quality of Education of National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine, *maria@perestyuk.com*. Translated from Kibernetyka ta Systemnyi Analiz, No. 2, March–April, 2021, pp. 69–80. Original article submitted November 2, 2020.
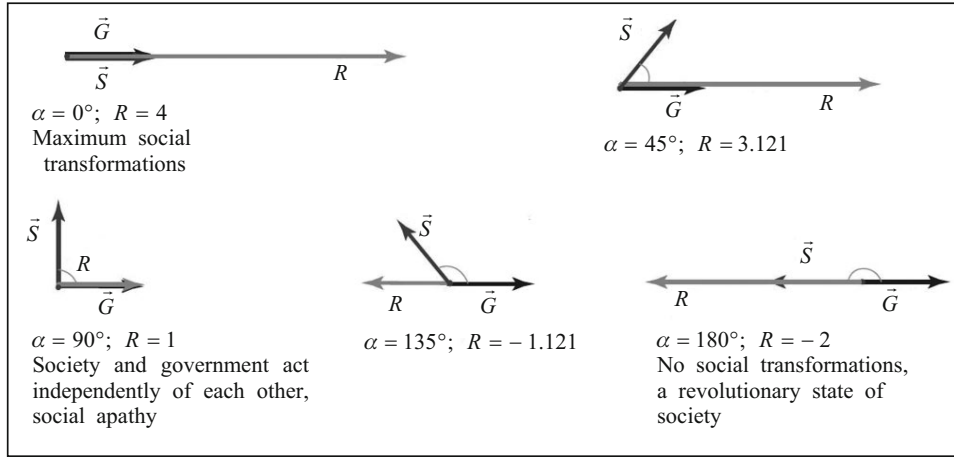
Fig. 1. The efficiency of social transformations (scope of reforms) for $K_S = 0.5$.

additional social activity (social energy) formed as a result of the interaction of citizens and is positive when society realizes that success generates new success and negative, when failures lead to loss of hope for positive change (apathy) and, consequently, generate new failures.

Let α be the angle between vectors $\vec{S}$ and $\vec{G}$ and K_S be the coefficient of the synergy of society. For different states of society, which can vary from social uplift and optimism to fatigue and apathy, the coefficient of synergy of society K_S can take on different values. In this study, we consider the following values of the coefficient of the synergy of the society: $K_S = [0.25; 0.5; 0.75; 1.0]$.

Usually, the dependence of the synergy of society on the angle α is nonlinear. The boundary values of the angle α are as follows: $\alpha = 0°$, when these two vector directions coincide (maximum social transformations and maximum positive synergy of society), and $\alpha = 180°$, when their directions are opposite (no social transformations, the maximum negative social energy is formed, which significantly exceeds the energy of government actions, resulting in a revolutionary state of society). In the case of $\alpha = 90°$, the vector of government actions and the vector of social activity (expectations) of society are orthogonal. This means that for the $\alpha = 90°$ government and society will act independently of each other.

To determine the social transformation factor R that measures the scale of social transformations, we use the following model proposed in [1]:

$$R = |\vec{G}| + |\vec{S}| \cos \alpha + K_S (|\vec{G}| + |\vec{S}|)^2 \cos \alpha; \ \alpha \in [0°, 180°].$$ (1)

In this model, the angle α is estimated (with the subsequent calculation of the social transformation factor R) both using the method of expert assessments (Delphi method) and by automatic online monitoring of public sentiment in the information space, or by both methods at once.

ASSESSMENT OF THE SCALE OF POTENTIAL SOCIAL TRANSFORMATIONS USING THE DELFI METHOD

The Miller scale (1,..,7) is used to determine the expert estimates of the angle α in the Delphi method. Score 7 of the Miller scale corresponds to the full support of society for government action ($\alpha = 0°$), and score 1 indicates that society does not support fully the actions of the authorities ($\alpha = 180°$). Let us choose the sampling step 0.5 of the Miller scale. We transform the score of the Miller scale into the value of the angle α and use model (1) to calculate the social transformation factor R for different values of the society synergy coefficient K_S (Table 1).

TABLE 1. Assessment of Social Transformations Efficiency

Miller Scale	Deviation Angle α (°)	Social Transformation Factor R			
		$K_S = 0.25$	$K_S = 0.5$	$K_S = 0.75$	$K_S = 1.0$
7	0	3.000	4.000	5.000	6.000
6.5	15	2.932	3.898	4.864	5.830
6	30	2.732	3.598	4.464	5.330
5.5	45	2.414	3.121	3.828	4.536
5	60	2.000	2.500	3.000	3.500
4.5	75	1.518	1.776	2.035	2.294
4	90	1.000	1.000	1.000	1.000
3.5	105	0.482	0.224	-0.035	-0.294
3	120	0.000	-0.500	-1.000	-1.500
2.5	135	-0.414	-1.121	-1.828	-2.536
2	150	-0.732	-1.598	-2.464	-3.330
1.5	165	-0.932	-1.898	-2.864	-3.830
1	180	-1.000	-2.000	-3.000	-4.000

To summarize individual expert opinions and obtain a collective expert assessment, we use a weighted mean

$$E = \sum_{i=1}^{n} w_i \cdot e_i, \tag{2}$$

where E is a collective expert estimate; w_i, $i = \overline{1, n}$, $\sum_{i=1}^{n} w_i = 1$, is the weight of individual estimate of the ith expert; and e_i, $i = \overline{1, n}$, is the individual estimate of the ith expert.

The weights w_i, $i = \overline{1, n}$, should be proportional to the deviation of the individual expert estimate from the average value of all individual expert estimates; therefore, they are calculated by the following formulas:

$$d_i = |\overline{E} - e'_i|; \quad e'_i = \frac{e_i - M_{\min}}{M_{\max} - M_{\min}}; \quad \overline{E} = \frac{\sum_{i=1}^{n} e'_i}{n};$$

$$w_i = \frac{1 - d_i}{\sum_{i=1}^{n}(1 - d_i)}, \tag{3}$$

where d_i, $i = \overline{1, n}$, is the distance of the normalized individual estimate e'_i of the ith expert from the average of a group of experts estimate $\overline{E}$ and $M_{\min} = 1$ and $M_{\max} = 7$ are the maximum and minimum scores of the Miller scale, respectively; other notation is the same as in the above formulas.

Since the Miller scale is an ordinal one, to estimate the consistency of collective expert estimate C, we use the weighted aggregate distance of individual expert estimates from one obtained by formula (2):

$$C = \frac{\sum_{i=1}^{n} |E - e'_i|}{C_{\max}}; \quad C_{\max} = n(n-1)(M_{\max} - M_{\min}). \tag{4}$$

The value $C = 1$ corresponds to a fully consistent estimate (when all experts give the same estimate), the value $C = 0$ corresponds to a situation where the set of obtained individual estimates contains the same number of minimum ($M_{\min}$) and maximum ($M_{\max}$) scores on the Miller scale, which indicates a lack of consensus in the expert group and the impossibility of obtaining a consistent collective expert opinion.

In the case where a generalized collective expert opinion on k individual issues is needed, the weighted sum

$$\widetilde{E} = \sum_{i=1}^{k} w_i' E_i; \quad w_i' = \frac{C_i}{\sum_{j=1}^{k} C_j}, \tag{5}$$

can be used as such an estimate, where $\widetilde{E}$ is the collective expert estimate for the set of k questions, E_i is a collective expert estimate for ith question, calculated by formula (2), w_i', $\sum_{i=1}^{k} w_i' = 1$, is the weighting factor of estimate E_i, which is proportional to its consistency C_i, calculated by formula (4).

In this case, the consistency of the estimate $\widetilde{E}$ can be determined as the average value

$$\widetilde{C} = \frac{\sum_{i=1}^{k} C_i}{k}. \tag{6}$$

Expert surveying with the proposed approach to assessing the scale of social changes involves the formation of an expert group, providing experts with a questionnaire containing questions on specific government actions over a certain time period, obtaining individual estimates of the angle α by Miller scale from experts, and calculating their consistency by formulas (2)–(6), along with the interpretation of the results.

AUTOMATED ONLINE ASSESSMENT OF THE SCALE OF POSSIBLE SOCIAL TRANSFORMATIONS

For automatic online estimating angle α (as a measure of inconsistency between government actions and society's expectations) along with the social transformation factor R, we use the method of the sentiment analysis of information messages obtained by online monitoring of relevant media resources and social networks.

Statistically, social networks accumulate 30–40% of information related to network users' reaction to important social events and public policies. The specificity of this information, namely its unstructured nature, low level of persistence, and the presence of fake messages, necessitates the development of specialized information systems whose main purpose is the collection, storage, and analytical processing of unstructured data obtained from online media and social networks.

The concept of simultaneous application of methods and means of information retrieval, analysis, and aggregation of data from modern information streams, which is the basis of such systems, allows scanning and initial processing information from websites, social networks, messengers, etc., automatically, create full-text databases, identify similar-content information messages, extract concepts and analyze a linguistic content of text messages, determine the tone of messages, and study the dynamics of thematic information flows. One of the main points of this paper is the significance given to identifying the sentiment of information in the process of content analysis of publications [2]. Sentiment polarity, as an emotional attitude of statement's author to an object, expressed in the text, in general can be defined as a function (in the simplest case, the sum) of sentiment polarities of its constituent units (sentences) and the rules of their combination [3]. We propose to use the method of determining the sentiment polarity of information by applying statistical patterns associated with the presence of certain lexical units in the texts, the naive Bayesian approach, and the method of neural networks [4].

The proposed approach to quantifying the scale of social transformations R is based on the hypothesis of adequate reflection of the message sentiments on the components of a vector of public expectations $\vec{S}$.

Let us consider the vector of public expectations $\vec{S}$ as

$$\vec{S} = \vec{S}_- + \vec{S}_= + \vec{S}_+,$$

where $\vec{S}_-$ is a component of negative attitude (actions), which corresponds to the opposition of the society to the actions of the authorities; $\vec{S}_=$ is a neutral component, which corresponds to the neutral attitude of the society to the actions of the authorities; $\vec{S}_+$ is a positive component, which corresponds to the approval (support) of government's actions by the

society. According to this definition, vectors $\vec{S}_-$ and $\vec{S}_+$ are collinear to the vector $\vec{G}$ of government's actions (the directions of vectors $\vec{S}_+$ and $\vec{G}$ coinside and the directions of vectors $\vec{S}_-$ and $\vec{G}$ are opposite), and vector $\vec{S}_=$ is orthogonal to vector $\vec{G}$. If we choose an orthogonal basis in such a way that the condition $\vec{G} = (0, 1)$ is satisfied, then the following relations hold:

$$\vec{S}_- = (-|\vec{S}_-|, 0), \ \vec{S}_= = (0, |\vec{S}_=|), \ \vec{S}_+ = (|\vec{S}_+|, 0),$$

$$\vec{S} = (|\vec{S}_+| - |\vec{S}_-|, |\vec{S}_=|),$$

$$|\vec{S}| = \sqrt{(|\vec{S}_+| - |\vec{S}_-|)^2 + |\vec{S}_=|^2},$$

$$\cos\alpha = \frac{|\vec{S}_+| - |\vec{S}_-|}{|\vec{S}|}, \ \alpha \in [0°, 180°].$$

(7)

Assume that the social activity of the society is reflected in the flow of messages of various media platforms, and the sentiment of these messages corresponds to the attitude (actions) of individuals and/or social groups to certain events, including those that are a consequence of government activity.

Let a certain action of the authorities $\vec{G}$ cause a sample of messages

$$M = M_- \cup M_= \cup M_+, \ M_i \cap M_j = \varnothing, \ i, j \in \{-, =, +\}, \ i \neq j,$$

where $\{-, =, +\}$ are negative, neutral, and positive sentiment polarities of messages. Then the following relations hold:

$$|\vec{S}_-| \sim |M_-| \sim p_- = \frac{|M_-|}{|M|},$$

$$|\vec{S}_=| \sim |M_=| \sim p_= = \frac{|M_=|}{|M|},$$

(8)

$$|\vec{S}_+| \sim |M_+| \sim p_+ = \frac{|M_+|}{|M|},$$

where $p_i, i \in \{-, =, +\}$, is the portion of messages with negative, neutral, and positive sentiment polarity, respectively.

With allowance for (4), we obtain an estimate

$$|\vec{S}| \sim \sqrt{(p_+ - p_-)^2 + p_=^2}.$$

(9)

We will use formulas (7)–(9) in model (1), and the resulting quantitative estimate of the social transformation factor R, reduced to the range (1,..., 7) according to the Miller scale (see Table 1), for verifying the consistency of quantitative and expert estimates, i.e., a social transformation factor R (scope of reforms) and the angle α (as a measure of inconsistency between the government actions and expectations of the society).

Let us consider the cases $p_+ = p_-$, $\cos\alpha = 0$. If $p_= = 1$, then the entire set of messages has neutral polarity. If $p_+ = p_- = 0.5$, $p_= = 0$, then we observe the maximum polarization of opinions in the society regarding government actions, but in both cases $R = 1$.

In order to distinguish these states, we define the conflict coefficient as

$$C = -(p_+ + p_-)\frac{c_+ \log c_+ + c_- \log c_-}{\log 2},$$

where $c_+ = p_+ / (p_+ + p_-)$ and $c_- = p_- / (p_+ + p_-)$.

The conflict coefficient C is a weighted entropy and varies in the range $[0,1]$. The value of $C = 1$ corresponds to the states $p_+ = p_- = 0.5$ and $p_= = 0$.
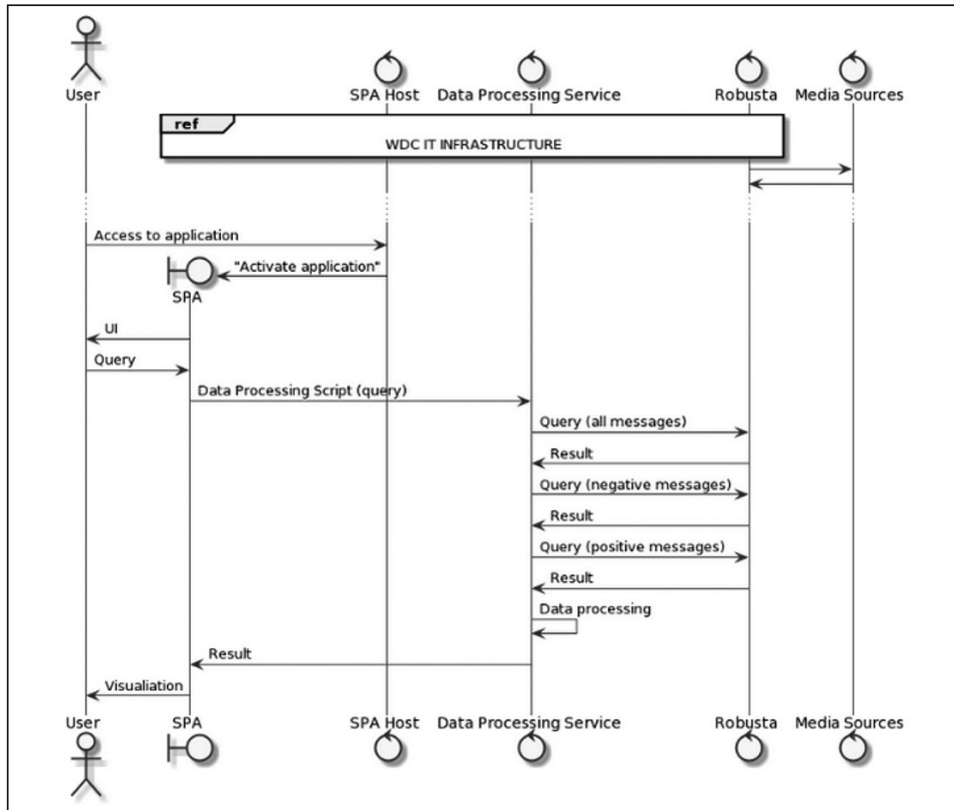
Fig. 2. Diagram of interaction of components of the World Data Center "Geoinformatics and Sustainable Development" Advanced Analytics online platform when processing SPA "PRO ET CONTRA" requests.

ADVANCED ANALYTICS ONLINE PLATFORM FOR ASSESSMENT OF SOCIETY'S ATTITUDE TOWARD GOVERNMENT ACTIONS

The proposed methods were used to implement a specialized application of the Social BI class [5], developed as a dashboard using tools and services of the online platform Advanced Analytics of the World Data Center "Geoinformatics and Sustainable Development" [6].

This integration platform is a distributed information system that includes services for the development and publication of single-page applications (SPA), data management and processing services, automatic bigdata collection, and mining services. The peculiarity of this system is the use of integration tools developed by the World Data Center "Geoinformatics and Sustainable Development" (WSD-Ukraine) that allows combining a variety of heterogeneous resources and services within a single IT infrastructure [7, 8].

The scheme of interaction of the components of the online platform is shown in Fig. 2.

In the process of implementing the application to assess society's attitude to government action based on the analysis of data from online media and social networks (SPA "PRO ET CONTRA"), the following components of the online platform were used: a system for developing and publishing single-page applications (SPA Host); Data Processing Service; a system of monitoring of social media (Media Sources), and linguistic analysis of bigdata obtained from social media (Robusta).

APPLICATION OF THE ADVANCED ANALYTICS ONLINE PLATFORM IN THE CONTEXT OF COVID-19 SPREAD

In what follows, we provide an analysis of public attitude to government action to strengthen quarantine measures in view of the spread of the COVID-19 pandemic. These results were obtained using the "PRO ET CONTRA"
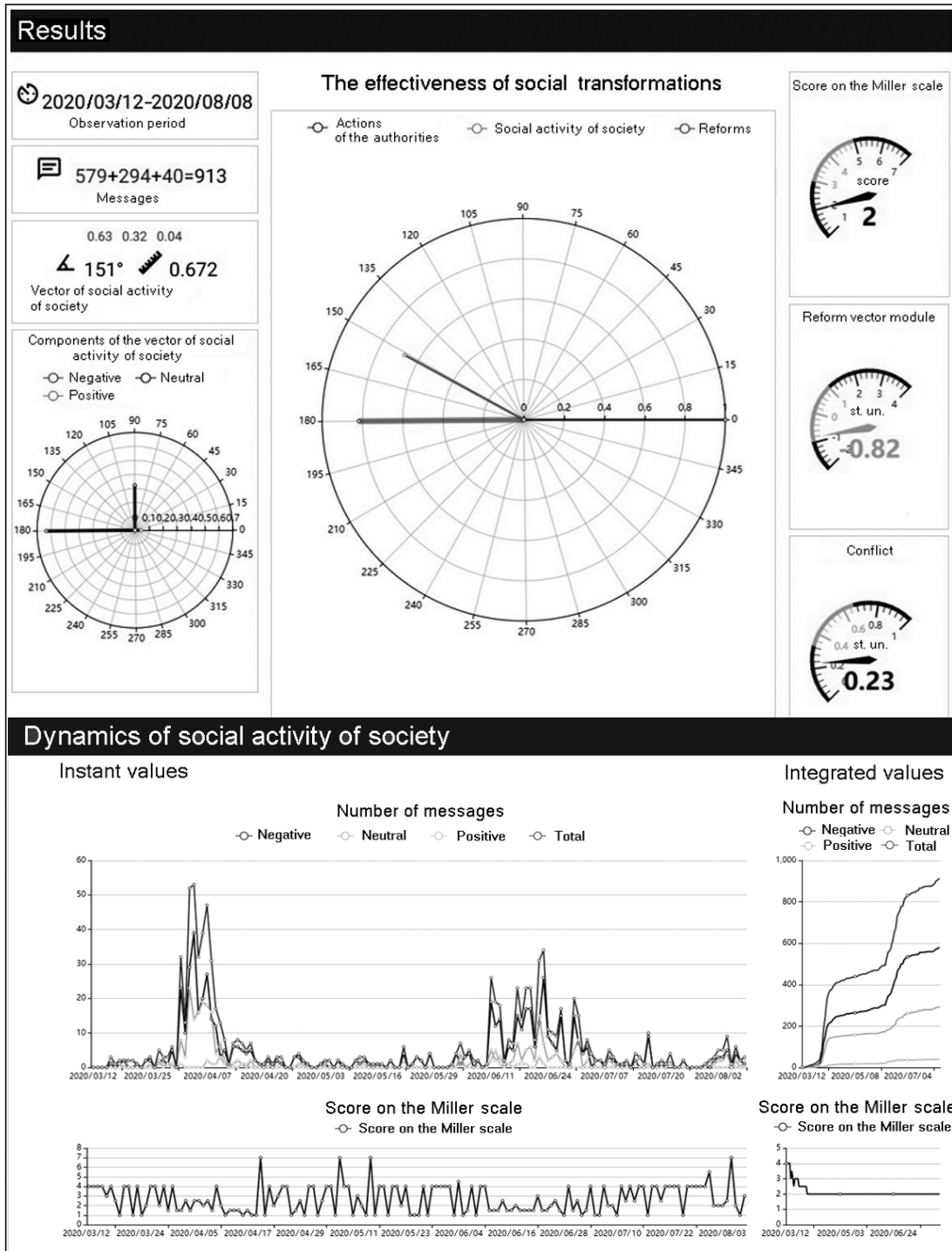
Fig. 3. Dashboards of "PRO ET CONTRA application.

application for the test data set of the Robusta system, accumulated over 200 days from the most popular 11 social media and 100 Ukrainian news websites (sample size is more than 2 million messages).

The query (\"ужесточен карантин\"|\"посилен карантин\") (україн|украин), (\"quarantine is escalated\") (in Ukrainian), formulated in the Robusta system language, was used to select messages containing such phrases as "ужесточение карантина" or "посилення карантину" in different cases and plurals, as well as words that begin with "україн" or "украин" (not case sensitive).

Fig. 4. Fragment of the questionnaire of the expert survey (dashboard).

Figure 3 presents a thematic panel of the general results of the analysis, which shows that during the period from 2020/03/12 to 2020/08/08, 913 messages were selected. Of these, 579 are negative, 294 are neutral, and 40 are positive (or in decimal form 0.60, 0.32, and 0.04, respectively). The relationships between these types of messages are shown in the vector diagram (Fig. 3). Based on the data, the estimates of the norm of the vector of social activity of the society $|\vec{S}| = 0.672$ and $\alpha = 151°$ were calculated. The vector diagram in the center of the dashboard (Fig. 3) shows the relationship between the vectors of government action $\vec{G}$, social activity (expectations) of society $\vec{S}$, and social transformation factor R.

The results of the analysis show that the attitude of the society to the actions of the authorities aimed at strengthening quarantine measures is quite consolidated (conflict ratio $C = 0.23$) and substantially negative (2 points score on the Miller scale). This causes resistance to the actions of the authorities ($R = -0.82$) in the form of ignoring quarantine restrictions.

As can be seen from Fig. 3, we have two peaks of revival of social activity of the society. The first peak (2020/04/01–2020/04/10) is associated with the declaration of emergency state and restrictive quarantine measures. On March 25, a state of emergency was declared throughout Ukraine; quarantine extended until April 24, trade establishments closed, except for grocery stores, pharmacies, gas stations, and banks. Metro transportation in Kyiv, Kharkiv, and Dnipro have been stopped; intercity and interregional automobile, railway, and air traffic have been suspended. On April 2, the government adopted a decree to strengthen quarantine restrictions. On April 6, the following new restrictive quarantine measures were introduced in Ukraine: a ban on visiting parks and recreation areas, mandatory wearing of face masks in public places, and people over 60 were recommended to stay at home.

The second peak (2020/06/11–2020/06/30) is associated with a significant deterioration of the epidemic situation in Ukraine and the continuation of adaptive quarantine. Thus, on June 11, the Ministry of Health reported to quarantine escalation. In particular, restrictions on entry to Ukraine for foreigners were tightened (insurance policy is needed). On June 12, there was information on the Government's plans to abolish the principal external independent testing (EIE) for applicants entering higher education institutions in 2020. Later, the Ministry of Health published the following three scenarios for conducting the EIE in 2020: to conduct this test within the predetermined period, from June 25 to July 17, change the dates, or cancel the test at all. On June 17, the Government decided to extend the adaptive quarantine until July 31 with updated rules for its observance. On June 22, mass inspections were initiated by the National Police for compliance with quarantine rules in stores, catering establishments, consumer services, and transport facilities. New criteria for easing/escalating quarantine measures in the regions have also come into force.

As a result of the analysis of the government actions during the second peak of activity, a questionnaire of four questions was formed, whose fragment is shown in Fig. 4.
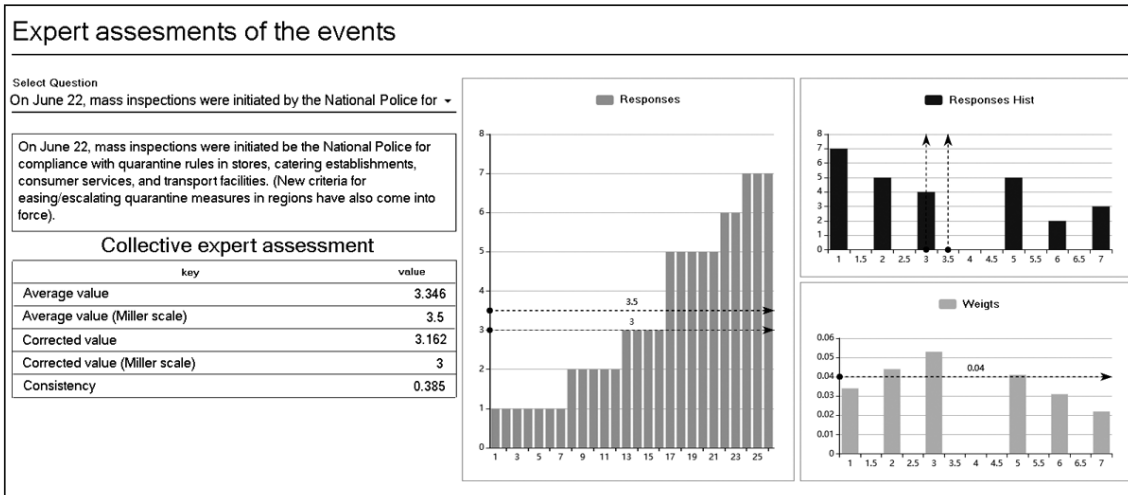
Fig. 5. Dashboard for the analysis of collective expert assessment.

TABLE 2. Results of the Expert Survey

Government Actions	The result of Expert Assessment of Society's Attitude to Government Action (on the Miller Scale)	Consistency of Expert Assessments
On June 11, due to a significant deterioration of the epidemic situation in Ukraine, the Ministry of Health resorted to quarantine escalation. In particular, restrictions on entry to Ukraine for foreigners were tightened (insurance policy was needed).	4	0.35
On June 12, there was information on the Government's plans to abolish the principal external independent evaluation (EIE) for applicants entering higher education institutions in 2020. Later, the Ministry of Health published the following three scenarios for conducting the EIE in 2020: to conduct this test within the predetermined period (from June 25 to July 17), to change the dates, or cancel the test at all.	3	0.49
On June 17, the Government decided to extend the adaptive quarantine until July 31 with updated rules for its observance.	3.5	0.37
On June 22, mass inspections were initiated by the National Police for compliance with quarantine rules in stores, catering establishments, consumer services, and transport facilities. New criteria for easing/escalating quarantine measures in the regions have also come into force.	3	0.39
Questionnaire in total	3.34	0.40

This questionnaire was proposed to the group of 30 experts, who gave individual expert assessments of society's attitude to government actions. Based on these assessments, a collective expert assessment was formed and analyzed by an appropriate dashboard (Fig. 5).

The results of the expert survey are summarized in Table 2.

For the same period, the quantitative score on the Miller scale is 2 and the consistency of the estimate is 0.65.

Thus, the expert group gave more optimistic but less consistent estimates compared to the results of the message analysis.

CONCLUSIONS

1. A method for assessing the effectiveness of social transformations has been developed, based on determining the degree of inconsistency of government actions and expectations of the society and synergy (social activity) of people, depending on the specified degree of inconsistency.

2. Methods of expert and automatic online assessment of the scale of social transformations based on the results of expert surveys according to the Delphi method and the results of monitoring open online publications and social networks using linguistic analysis of a large number of messages with different sentiments are proposed.

3. The specialized application "PRO ET CONTRA" is designed based on the integration of various heterogeneous resources and services within the unified IT infrastructure of the World Data Center "Geoinformatics and Sustainable Development," which is a part of the software of the Information and Analytical Situation Center WDC-Ukraine and serves for expert and automated online assessment of the scope of social transformations.

4. The effectiveness of the proposed technique and online tools developed was demonstrated on the example of a quantitative assessment of the attitude of the population of Ukraine to the actions of the authorities aimed at strengthening quarantine measures caused by the spread of the COVID-19 pandemic.

REFERENCES

1. M. Zgurovsky, A. Boldak, K. Yefremov et al. Foresight and Construction of the Strategies of Socio-Economic Development of Ukraine on Mid-Term (up to 2020) and Long-Term (up to 2030) Time Horizons, Scientific advisor of the project acad. of NAS of Ukraine M. Zgurovsky, 2nd ed., NTUU "Igor Sikorsky KPI," Publ. House "Polytechnica," Kyiv (2016). URL: http://wdc.org.ua/en/node/186284.

2. F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, Sentiment Analysis in Social Networks, Morgan Kaufmann (2017).

3. A. G. Dodonov, D. V. Lande, and V. G. Putyatin, Computer Networks and Analytical Researches [in Russian], IPRI NAS of Ukraine, Kyiv (2014).

4. V. D. Lande, Fundamentals of the Integration of Information Flows [in Russian], Inzhiniring, Kyiv (2006)

5. Application "PRO ET CONTRA." URL: https://jace-dev.herokuapp.com/design/proetcontra#/.

6. Online platform Advanced Analytics, WDC-Ukraine. URL: http://open.wdc.org.ua/#/.

7. D. Korenko and A. Boldak, "Approach to organization of client-server interaction for implementation of model-view-controller pattern in distributed systems," in: Proc. International Conference on Security, Fault Tolerance, Intelligence (ICSFTI2019) (Kyiv, Ukraine, 14–15 May, 2019), Igor Sikorsky Kyiv Polytechnic Institute, Publ. House "Polytechnica," Kyiv (2019), pp. 182–187.

8. A. Boldak and K. Yefremov, "WDC-Ukraine's distributed platform for supporting research data life cycle," in: Proc. 17th Intern. Conf. on System Analysis and Information Technologies (SAIT 2015) (Kyiv, Ukraine, 22–25 June, 2015), Kyiv (2015), pp. 210–211.