

# COVID-19 MEASURES SENTIMENT ANALYSIS BASED ON A SOCIAL NETWORK DATASET

Kateryna Kononova<sup>1</sup>, Rostyslav Lutsenko<sup>2</sup>

<sup>1</sup>D.Sc., professor, Karazin Kharkiv National University, e-mail: [kateryna.kononova@karazin.ua](mailto:kateryna.kononova@karazin.ua)

<sup>2</sup>M.S., Karazin Kharkiv National University, e-mail: [roxanisen@gmail.com](mailto:roxanisen@gmail.com)

**Abstract.** The paper discusses the prototype of a public opinion monitoring system toward Covid-19 measures using machine learning and sentiment analysis, which includes: 1) data collection; 2) data preprocessing and statistical analysis of a created corpus; 3) semantic analysis of the corpus; 4) sentiment analysis of the data. The sample of posts, which was collected for the period April-May 2020 using 360 unique search queries, includes 6726 publications of Ukrainian Facebook users. The assessment obtained by the proposed methodology is confirmed by the results of the survey about supporting the government's Covid activities, according to which only about 10% of respondents are positive about the government's actions, and more than 60% are negative. Thus, the proposed system, developed as a set of Python and SQL scripts, can be recommended for implementation.

**Keywords:** coronavirus, government, public, survey, machine learning, semantic analysis, sentiment analysis.

## 1. Introduction

The SARS-COV-2 pandemic has been going on already for more than a year. It is characterized by rapid spread, high mortality, and lack of specific treatment. The pandemic has become a factor that forced the governments of many countries to reconsider economic and social policies and formulate new development priorities.

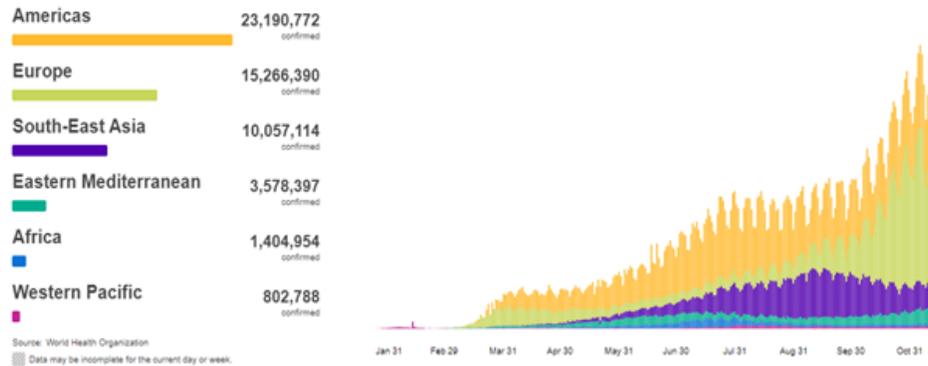
The disease, which started in December 2019 in Wuhan (Hubei Province, China), was recognized by the WHO as a pandemic on March 11, 2020 (World Health Organization, 2020). Now the pandemic spread to almost all countries (Fig. 1). After the Covid outbreak in March-April 2020, the number of infected reached the plateau. However, in late August, the second wave of the disease began.

In Ukraine, the first case of the coronavirus was registered in the Chernivtsi region on March 3, 2020; the active phase of the pandemic began on March 25 (Public Health Center of the Ministry of Health of Ukraine, 2020). The introduction of a lockdown in March-April delayed the spread of the coronavirus. However, after the weakening of the quarantine regime, the number of patients started growing again (World Data Center for Geoinformatics and Sustainable Development, 2020). In May-July 2020, there was some stabilization, but just after the summer season, the number of infected people began to rise

---

*Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).*

sharply. In autumn-winter 2020, the epidemical situation in Ukraine deteriorated, leading to the second lockdown in January 2021 (World Center data on geoinformatics and sustainable development, 2021).



**Fig. 1.** COVID-19 dynamics (World Health Organization, 2020)

Covid-19 has an extremely negative impact on the global economy: according to the World Bank, a global GDP contraction of 5.2% was expected in 2020 (Table 1). The income per capita might fall 3.6%, leaving millions of people in poverty.

**Table 1.** GDP forecast<sup>1</sup> (World Bank, 2020)

	2018	2019 e <sup>2</sup>	2020 f <sup>3</sup>	2021 f
World	3.0	2.4	-5.2	4.2
Advanced economies	2.1	1.6	-7.0	3.9
Developing economies	4.3	3.5	-2.5	4.6

Governments are eager to find and implement proper actions to combat the coronavirus spread, as well as to overcome the negative economic consequences of the pandemic. The key to success in overcoming the pandemic is the consolidation of trust between governments and people: the effectiveness of Covid-19 control measures depends on the conscious attitude and responsibility of each citizen.

This work aims to develop a prototype of a public opinion monitoring system toward Covid-19 measures. This could be done in a form of a regular survey or by applying machine learning and sentiment analysis to the data retrieved from social media.

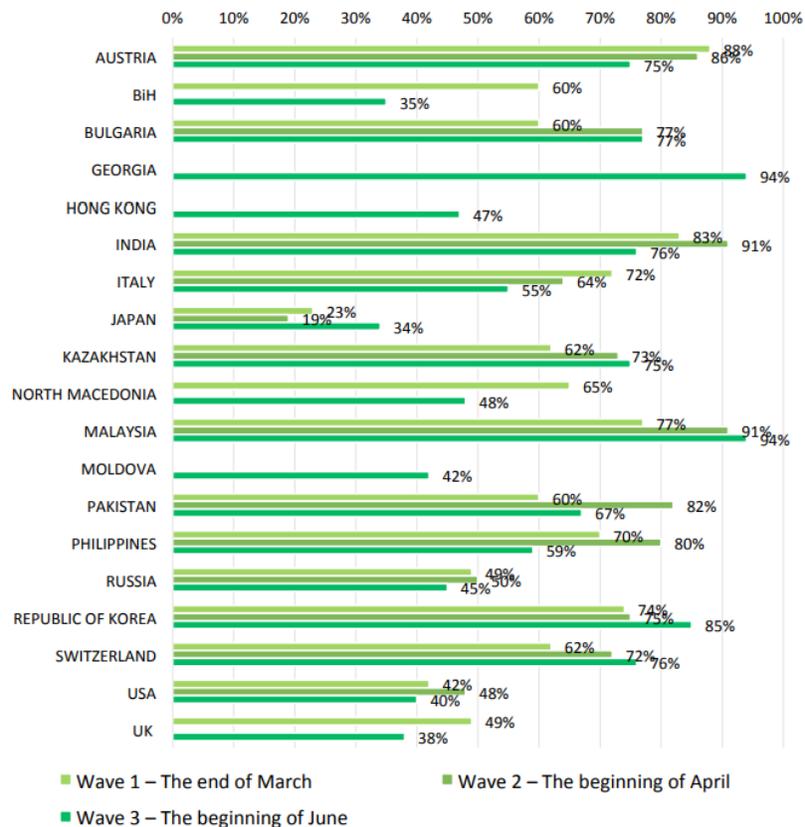
<sup>1</sup> percent change from the previous year

<sup>2</sup> estimate

<sup>3</sup> forecast

## 2. A survey approach

In a pandemic, Gallup International experts have studied the issue of the government's actions public support<sup>4</sup>. According to a three-wave survey in June 2020, the highest support<sup>5</sup> was observed in Georgia (94%), Malaysia (94%), and the Republic of Korea (85%). The lowest levels were in Japan (34%), Bosnia and Herzegovina (35%), the United Kingdom (38%), and the United States (40%). The results in the dynamics are also quite interesting (Fig. 2): the data shows that support may increase (Kazakhstan, Malaysia, Republic of Korea, and Switzerland), decrease (Austria, Italy, Macedonia, and the United Kingdom), or fluctuate (India, Pakistan, Philippines, Russia, and the United States).

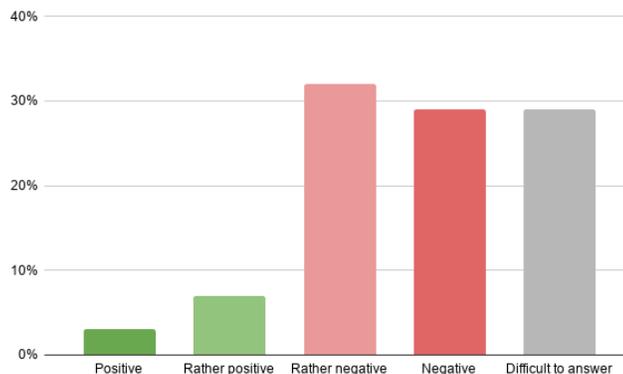


**Fig. 2.** Support of the government's anti-epidemic measures (Gallup International, 2020)

<sup>4</sup> The question was: to what extent do you agree or disagree with the following statement – 'I think the government is doing well with the coronavirus'

<sup>5</sup> Strongly agree and agree with the statement above

Unfortunately, Ukraine has not participated in this survey. To measure Ukrainians' sentiments towards the Covid measures, we conducted an online survey<sup>6</sup>, in which 1600 respondents took part. The results showed that only about 10% of Ukrainians are positive about the government's actions, while more than 60% are negative about them (Fig. 3).



**Fig. 3.** What is your attitude toward the Covid measures?

### 3. A sentiment analysis approach

To develop a prototype of a public opinion monitoring system toward Covid-19 measures we proposed applying machine learning and sentiment analysis to the content of the social network<sup>7</sup>. As a platform for the analysis, Facebook was chosen, because this social network is the largest social media used by more than two billion people every month. According to Research & Branding Group, the number of active users in Ukraine is about 63% of the population, which is approximately 20.8 million people (FutureNow, 2020). Moreover, Facebook has features that could enrich the results of the semantic analysis: a quantity of 'likes' allows assessing the content relevance, the types of reactions show user's attitude to the post, the number of reposts reflects its distribution among other users.

The development of the system includes the following steps, which are implemented in the form of SQL and Python scripts:

- I. Data collection.
- II. Data preprocessing and statistical analysis of a created corpus.
- III. Semantic analysis of the corpus.
- IV. Sentiment analysis of the data.

<sup>6</sup> Poll link (<https://docs.google.com/forms/d/e/1FAIpQLSe3pk8AONnLr6PB1VJCBAHWzF7tBhnS2wk8VKkEoL-SVHiwDQ/viewform>) was posted on Telegram channels

<sup>7</sup> According to the Institute of Sociology NAS of Ukraine, most Ukrainians receive information about the pandemic from the media and social networks

At the first stage, raw data was downloaded from Facebook using the DATA365 API. To identify posts discussing the Covid measures, more than 360 unique search queries have been created. The sample, which was collected for the period April-May 2020, includes 6726 publications of Ukrainian users (Fig. 4).

id	created_time	text	lang	tags	reaction	comment	share	like	love	haha	wow	sad	angry
577034592905048	2020-04-01 15:00:54	Поки у Верховній Раді тривають...	uk	(tstypin_информа...	12	0	0	12	0	0	0	0	0
1134005400283184	2020-04-01 16:39:19	ПАНДЕМІЯ ТА ВИПЛАТИ 🇺🇦 Як де...	uk	(versustawgroup)	6	0	0	6	0	0	0	0	0
609418609785175	2020-04-04 09:45:02	На Тернопільщині - спалах корон...	uk	(tokshoy_ukraine...	13	1	4	4	0	0	2	7	0
250565683081273	2020-04-05 09:40:00	Якщо розповсюдження хвороби н...	uk	(03ombr_00c0001...	486	15	76	471	15	0	0	0	0
2584823391733168	2020-04-06 08:57:09	Сидіть вдома і дивіться телевізор...	uk	(zelenskyy)	8	6	0	6	0	2	0	0	0
613301306063572	2020-04-09 14:22:50	Пандемія чи безробіття? Четвер, 9...	uk	(viv_zelenskyy)...	8	5	2	7	0	0	0	0	1
2343500715948370	2020-04-14 13:41:13	#хронікарентину Визиває інте...	uk	(хронікарентин)...	8	3	0	4	2	2	0	0	0
2624127084491339	2020-04-15 13:02:08	Зараз таке враження, що Україну ...	uk	(відповідь)ама...	9	2	2	9	0	0	0	0	0
62332215060481	2020-04-22 14:18:19	ЧИ ДОВГО УКРАЇНА БУДЕ "НА КАР...	uk	(govorytyvelyviv...	30	23	1	25	0	2	2	0	1
2680659402193324	2020-04-30 12:18:24	🇺🇦🇺🇦🇺🇦 Володимир Зеленський ...	uk	(covid_19_шви)...	5	0	1	2	0	0	1	0	2

Fig. 4. A fragment of the collected sample<sup>8</sup>

At the second stage, the texts were pre-processed:

1. Lowercase letters formatting to avoid repeating words;
2. Redundant characters removing (numbers, punctuation, etc.);
3. Stop words deleting (such as 'this', 'and', etc.);
4. Stemming (reducing words to the root form).

After the corpus had been created, the list of the most used words was analyzed (Fig.

5).

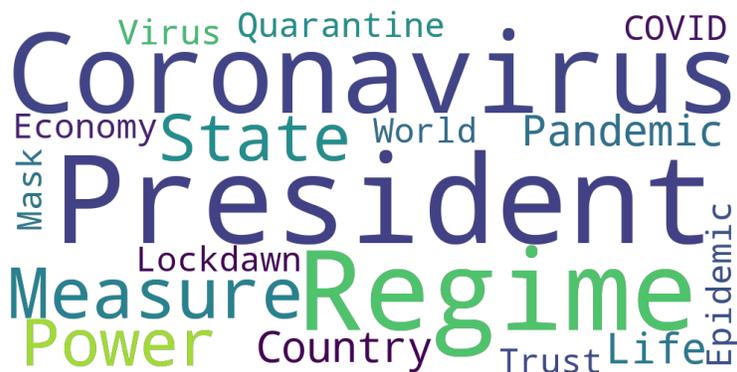


Fig. 5. Top words

Semantic analysis of the corpus (stage III) was performed using Word2Vec tools. The Word2Vec model is a two-layer neural network that learns to reconstruct the linguistic context of words. As a result, words that have a common context in the corpus are placed close to each other in the vector space. This makes it possible to analyze the main

<sup>8</sup> The data includes id, time, text, language, tags, reaction\_count, comment\_count, share\_count, reaction\_like\_count, reaction\_love\_count, reaction\_haha\_count, reaction\_wow\_count, reaction\_sad\_count, reaction\_angry\_count

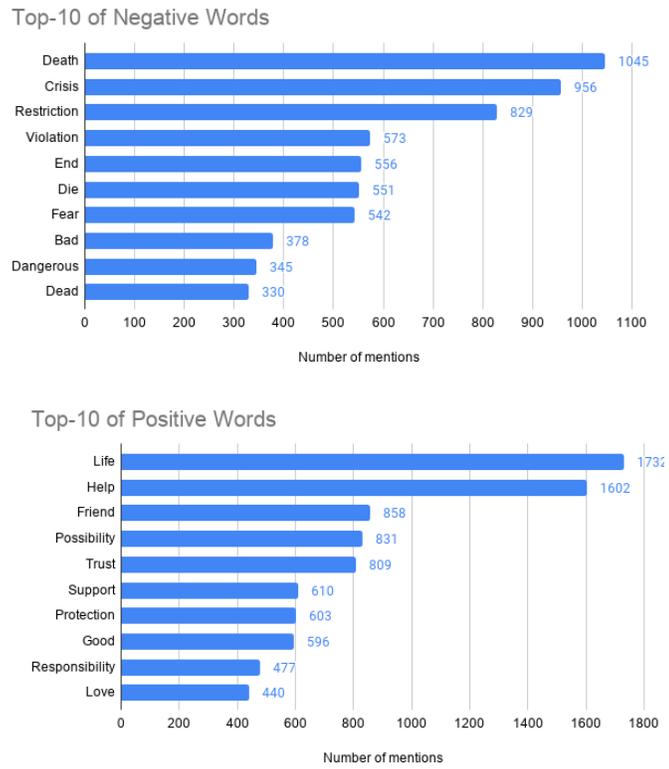
narratives. Using the Python word2vec library (Python word2vec Library, 2020), we calculated the distances between the words (Table 2).

**Table 2.** The most commonly used phrases

<b>CORONAVIR</b>		<b>LOCKDOWN</b>		<b>MASK</b>	
state budget	0.98	curfew	0.90	wear	0.88
epidemic	0.97	introducing	0.84	glove	0.86
complications	0.97	fortnightly	0.81	antiseptic	0.85
pandemic	0.96	regime	0.81	pharmacy	0.84
accident	0.96	violation	0.79	protective	0.83
outbreak	0.96	strengthening	0.77	respirator	0.83
measures	0.95	penalty	0.76	disposable	0.77
morbidity	0.84	severe	0.75	free	0.75
<b>GOVERNME</b>		<b>PRESIDENT</b>		<b>TRUST</b>	
resistance	0.99	Zelensky	0.94	influence	0.97
infographics	0.99	prime minister	0.85	society	0.97
ministry	0.99	to declare	0.85	understanding	0.97
continue	0.98	minister	0.80	to blow up	0.97
extraordinary	0.98	supreme	0.77	interest	0.96
attenuated	0.97			important	0.96
strengthen	0.97			idea	0.96

As we can see from Table 2, posts about the coronavirus most often discuss the state budget and the Covid-19 control measures; in the context of the lockdown people talk about its duration, penalties, and violations, when talking about the government, it is about pandemic resistance and infographics on anti-epidemical measures.

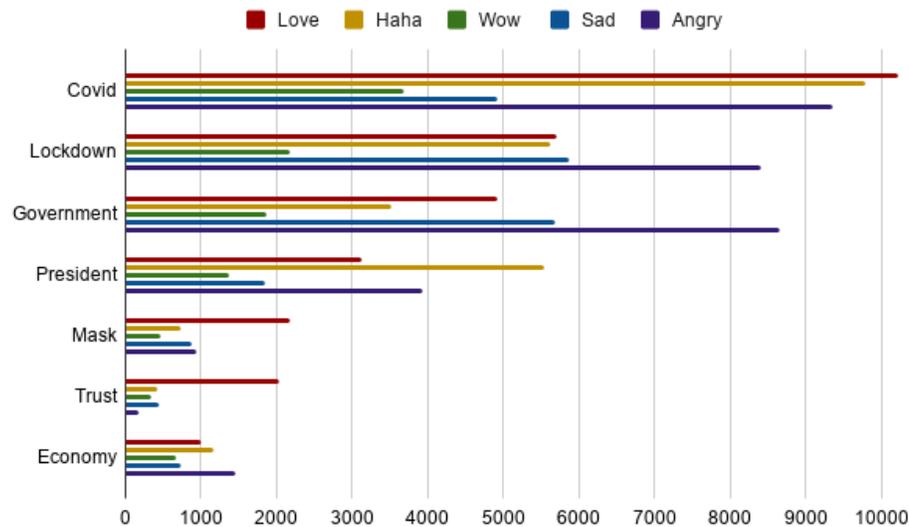
At the fourth stage, dictionaries of positive and negative words were constructed (Text Analysis API, 2020). In the positive dictionary (Fig. 6) one can see the words that are important for people in difficult times ('life', 'help', etc.). The words 'death', 'crisis', and 'restrictions' are on the top in the negative dictionary. Dictionaries comparing shows that on average, words with a positive connotation are used 30% more often than pessimistic ones. Even the words 'life' and 'death', which are at the top of both dictionaries, are correlated as 1.7 to 1.



**Fig. 6.** Top positive and negative words

Fig. 7 shows the total reactions of different types ('love', 'laughter', 'delight', 'sadness', 'anger'<sup>9</sup>) to the posts.





**Fig. 7.** Distribution of reactions of different types

Analysis of fig. 7 shows that the word ‘coronavirus’ causes very controversial emotions – ‘laughter’ and ‘anger’ are almost on the same level. The mention of the words ‘government’ and ‘lockdown’ often causes ‘anger’ and ‘sadness’; ‘mask’ – ‘love’ and ‘anger’, ‘president’ and ‘economy’ – ‘laughter’ and ‘anger’. According to Plutchik’s wheel of emotions (Bunyak, 2017), we see that ‘anger’ in combination with other emotions causes complex negative sentiments: aggression, contempt, pity, and disapproval. In particular, the sentiments about the President and Lockdown = ‘laughter’ + ‘anger’ = ‘aggression’, the Government = ‘anger’ + ‘sadness’ = ‘contempt’.

These results are in line with the survey results that allow to recommend implementing the public opinion monitoring system toward Covid-19 measures based on the social network’s content using machine learning and sentiment analysis.

#### 4. Conclusions

The paper discusses the prototype of the public opinion monitoring system toward Covid-19 measures based on the social network’s content using machine learning and sentiment analysis, which includes: 1) Data collection; 2) Data preprocessing and statistical analysis of a created corpus; 3) Semantic analysis of the corpus; 4) Sentiment analysis of the data. The sample of posts, which was collected for the period April-May 2020 using 360 unique search queries, includes 6726 publications of Ukrainian Facebook users.

The most frequently used words of the retrieved corpus were ‘coronavirus’, ‘pandemic’, ‘lockdown’, ‘mask’, ‘power’, ‘state’, ‘president’. Semantic analysis of the

corpus using Word2Vec tools showed that posts about coronavirus most often discussed the state budget and the Covid measures; in the context of the lockdown people talk about its duration, penalties and violations, when talking about the government, it is about pandemic resistance and infographics on anti-epidemic measures.

To analyze public sentiments, dictionaries of positive and negative words were created. Dictionaries comparing shows that on average, words with a positive connotation are used 30% more often than pessimistic ones. Analysis of reactions to the posts showed that the word ‘coronavirus’ causes very controversial emotions – ‘laughter’ and ‘anger’. At the same time, the mention of the words ‘government’ and ‘lockdown’ most often causes ‘anger’ and ‘sadness’, ‘president’ and ‘economy’ – ‘laughter’ and ‘anger’ (‘contempt’ and ‘aggression’ according to Plutchik’s methodology).

The assessment obtained by the proposed methodology is confirmed by the results of the survey about supporting the policy responses to Covid. According to this survey, only about 10% of respondents are positive about the government’s actions, and more than 60% are negative.

Thus, the proposed public opinion monitoring system toward Covid-19 measures based on the social network’s content using machine learning and sentiment analysis can be recommended for implementation.

## 5. References

1. Gallup International. (2020). Retrieved from <https://www.gallup-international.com/wp-content/uploads/2020/08/Covid-19Wave-3-1.pdf>
2. The world bank. (2020). Global economic prospects Retrieved from <https://www.worldbank.org/en/publication/global-economic-prospects>
3. World Health Organization. (2020). Retrieved from <https://Covid19.who.int/>
4. State Institution "Public Health Center of the Ministry of Health of Ukraine". (2020). Retrieved from <https://www.phc.org.ua/>
5. World Data Center for Geoinformatics and Sustainable Development. (2020). Retrieved from <http://wdc.org.ua/uk/Covid19-second-wave-growth>; <http://wdc.org.ua/uk/Covid19-transformation-after-pandemic-europe>
6. Pandemic COVID-19 and its consequences in the field of tourism in Ukraine. (2020). Retrieved from <http://www.ntoukraine.org/assets/files/EBRD-COVID19-Report-UKR.pdf>
7. Kyiv International Institute of Sociology. (2020). Sociological and marketing research Retrieved from <http://kiis.com.ua/?lang=ukr&cat=reports&id=964&page=3>
8. FutureNow. (2020). Future is here! The future is already with us! Retrieved from <https://futurenow.com.ua/21-najpopulyarnishyh-sotsialnyh-merezh-svitu-ta-rejtyng-ukrayiny/>
9. Text Analysis API. (2020). Retrieved from <https://neticle.com/textanalysisapi/en/analysis-method>
10. Python word2vec libraries. (2020). Retrieved from <https://papers.nips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf>

11. Bunyak, N. (2017). General psychology: lectures (part II). Ternopil: published by TNTU. I. Pulyuya
12. Influence of COVID-19 and quarantine restrictions on the economy of Ukraine. (2020). Retrieved from <https://www.me.gov.ua/Documents/Download?id=bc5d2c61-1a7f-4ec7-8071-b996f2ad2b5a>
13. World Data Center for Geoinformatics and Sustainable Development. (2020). COVID-19 in Ukraine: analysis and forecast. Retrieved from <https://wdc-ukraine.maps.arcgis.com/apps/opsdashboard/index.html#/07b38b42264d4a2ea5a05d578e235559>
14. Official website of the Ministry of Economy, Trade and Agriculture of Ukraine. (2020). Retrieved from <https://www.me.gov.ua/Documents/List?lang=uk-UA&tag=Konsensus-prognoz>
15. Official information portal of the Cabinet of Ministers of Ukraine “Coronavirus in Ukraine”. (2020). Retrieved from <https://Covid19.gov.ua/>
16. Official site of the Ministry of Health of Ukraine. (2020). Retrieved from <https://moz.gov.ua/koronavirus-2019-ncov>
17. Official site of the National Health Service of Ukraine. (2020). Retrieved from <https://nszu.gov.ua/>