ISSN 0147-6882, Scientific and Technical Information Processing, 2021, Vol. 48, No. 2, pp. 78–86. © Allerton Press, Inc., 2021.
Russian Text © The Author(s), 2021, published in Nauchno-Tekhnicheskaya Informatsiya, Seriya 1: Organizatsiya i Metodika Informatsionnoi Raboty, 2021, No. 4, pp. 8–17.

Methodological Approach to Use of Web Content by Small Business

N. V. Dneprovskaya[a, c] and I. V. Shevtsova[b, *]

[a] *Plekhanov Russian University of Economics, Moscow, 117997 Russia*
[b] *Lomonosov Moscow State University, Moscow, 119991 Russia*
[c] *HSE University, Moscow, 101000 Russia*
**e-mail: Shevtsova@spa.msu.ru*

Received January 18, 2021

Abstract—Digitization encourages accumulation of web content, an important resource of economic activity. Despite high level of development of technology of work with web content, the need for significant expenses restricts its use by small businesses. Web content characterized by absence of structure, diversity of sources, and high speed of data flow, is included in the concept of "Big Data", efficient work with which requires access to financial, computing, and labor resources. The developed and tested methodological approach to use of web content, taking into account capabilities of small business, enables a specialist in any subject area to upload textual information, convert it into a database, and analyze it using widespread or public domain software.

Keywords: web content, text analysis automation, digital economy, small business, digitalization of society
DOI: 10.3103/S0147688221020040

INTRODUCTION

Digital data is becoming an important resource in the global economy, which ensures creation of new forms of economic value such as improved service quality or other significant effects [1–3]. In the digital domain, business can better understand its customers, forecast their needs and expectations due to processing of digital footprints left in the course of their online activity.

Intensive development and expansion of information technologies (ITs) allowed to simplify web content creation and increase the rate of its accumulation, and current state of IT infrastructure of the society enables business entities to use web content as a new resource in its activity. Efficiency of digital data use depends largely on its structure and tools for work therewith. Various tools, analytical platforms, and application software were created to work with structured data. Automated processing of structured data for sure served as beginning of the epoch of informatization and information society of the XX century [4]. Research and improvement of data analysis tools continues, now they are widely used by almost all organizations in financial accounting, planning of production, process, and project management activities.

The epoch of the World Wide Web radically changes the tools, techniques, and habits of content creation. Web 2.0 services opened ways for each Internet user to express publicly their opinions on goods and services, to share their feelings and emotions, and to tell about significant events in their lives. More than 75% of Russians use social networks and are able to publish a variety of content in the network: photos (Instagram), videos (YouTube, TikTok), text, and various multimedia (VK, Twitter) [5]. The content published and freely distributed in social media contains potentially valuable facts for business, and its processing enables making important conclusions and receiving new information. Methods for its free use in solving political, economic, or scientific problems are open, but complexity of processing of web content consists of the fact that it is unstructured, multi-format, and comes from various sources.

Although noncommercial users publish web content free of charge and freely distribute it in the digital environment, digital platform providers, being large corporations, are those who receive the most benefit from its creation, distribution, and processing. It is caused by large investments of big players in studies of social media, development, and use thereof [6]. To enable Russian enterprises, including small and medium ones, to use the advantages of digitalization in their economic activities, it is required to develop and implement methods of analyzing the web content, which, on the one hand, will not require expensive investments and large labor content of specialists, and, on the other hand, will ensure relevance and objectivity of results for use in the process of management decision making.

Researchers suggest a number of measures to support small and medium businesses in the context of

digital economy, among which is creation of favorable conditions for making business and private investors in Russia, including reduction of tax burden, involvement of graduates of foreign universities, and development of social infrastructure [7]. Suggestions on support of small businesses in view of the digitalization of the society are mostly general by their nature and are intended for governmental bodies to form entrepreneurial climate.

Digital economy will be established when digital resources and web content, in particular, enable a wide range of businesses to create innovations, and to form value in business. In the business literature the issues on digitalization opportunities use are often considered in connection with large companies [8]. At the same time, small business needs guidelines that reveal clear and simple mechanisms for work with digital data, available to the same extent as platform Uber for taxi drivers. Provision of access for small businesses to new digital resources for development thereof is becoming a relevant task. In the economic activity of small business, effective mechanisms are required for use of digitalization resources, corresponding to its technological and professional capabilities.

In developed countries, small businesses create 60% of GDP, while in Russia this value has just recently exceeded the level of 20% [9], however, within the framework of the National project[1], the task is to increase its contribution to the economy over 30%. Analysis of trends in development of small and individual entrepreneurship indicates its growing role in provision of sustainability of society due to creation of conditions for employment and establishment of new jobs [10].

According to the results of analysis of promising areas for development of small business in developed countries, a lag in Russian theoretical, research, and practice investigation studies takes place [11]. Economists raise the problem of asymmetry in digitalization, which is observed both on the global scale, and in various industries. The digital asymmetry, expressed in unequal access to digital data and technologies, leads to negative consequences for national economies and business entities [12, 13].

The purpose of this study is to suggest a methodological approach to analytical processing of web content, which will help a wide range of entities in small businesses to use digital data as an economic resource.

The following tasks were set in our study:

• to analyze web content as an economic resource for small and medium business,

• to formulate requirements to methods of processing of web content applicable in the small business environment,

• to suggest a method of using web content for solving economic tasks which includes technologies and automation algorithms available to small business,

• to carry out testing of the suggested method.

Solving of these tasks requires an interdisciplinary approach, which includes economic theory for studying the digitalization of the economy [14], state-of-the-art management in the area of knowledge management [15], cutting edge IT in the analysis methodology [16–18], and text abstracting [19, 20]. The methods and techniques included in the theory of knowledge management enable formalizing information and using it as a resource for business development. The theoretical basis of the study also includes overview of methods of test data automatic processing, work on annotation and abstracting of scientific and technical information, and works on issues of natural language analysis [21, 22]. Selection and adaptation of knowledge management techniques and text processing technologies, taking into account the IT available to small business, contributes to support of small business in the digital epoch. The information base of the study is made of data taken from official statistics (Federal State Statistics Service, 2020), research organizations (HSE University), industry associations, and business publications, as well as data sets obtained from social networks. The applied study methods include overview, comparison, economic analysis, business process modeling, and testing of developed practical issues of the methodological approach.

CONCEPTUALISATION OF WEB CONTENT FOR SMALL BUSINESS

The flow of web content published by users is the kind of digital resources extracted from the digitalization of the society. Digital means (tools) of communication provide freedom for public expression of their opinion and publishing of any content by Internet users. Before beginning of digitalization, it was possible to collect data on opinion of consumers or citizens only in the process of expensive marketing researches or opinion polls. In the Internet, in textual form, users share their opinions on goods, services, and companies. Such messages contain information on comparison of characteristics of any goods and services important for consumers, disadvantages and advantages thereof, and the expectations of consumers. Large participants of the digital market, such as Google, have learnt to extract general impressions that accompany descriptions of the organizations as summaries marked on Google Maps, using automatic analysis of users' opinions.

With the digitalization progress of the society, social media are gaining great importance in marketing and advertising, therefore nearly all companies, including small and medium ones, do not leave this communication channel without control activity.

[1] Small and Medium Entrepreneurship National project. https://futurerussia.gov.ru/maloe-i-srednee-predprinimatelstvo (in Russian).

Companies publish at least information about themselves, use Internet advertising, and respond to obtained opinions. Peculiarities of social networks and chats as a channel of communication with a consumer in the Internet were disclosed in a number of publications, including from the point of view of automation thereof [23, 24]. Monitoring of web content is actively used by the departments responsible for public relations (PR) and marketing [25, 26], as well as for managing of the company image and brands.

Social networks and other web 2.0 services may be used not only for interaction with customers, suppliers, contractors, but also as a strategic management tool for determining competitive advantages at the market and revealing the new needs. Web content becomes a valuable source of information for business development.

Web content is the data contained in web and web 2.0 services and is characterized by the following:

• low structuring, in general, by the following parameters only: source, author, date, text, and responds of other users (likes, comments, and views);

• disseminating on multiple sources in the network for communication and exchange of opinions, such as social networks and forums. Various groups of consumers prefer various resources;

• continuous inflow of new messages which is created as response to different events. While for some services the data published a year ago and being already off the agenda are of no importance, the course of events of response to the company action and development of needs and degree of satisfaction with products is of interest for consumer study;

• various conditions of access to the web content. There is often no direct access; user registration is required at least.

Sources of web content available for analysis may be external and internal.

External sources include social media (social networks, forums), blogs (messages and comments to them), electronic media, specialized web platforms, for example, recruiting ones, where vacancies and resumes are published, wiki sites (specialized sites for professionals), and expert communities.

Internal sources include email and instant messaging services through which consumers directly send their questions and suggestions. Difference between business correspondence and external web content is that the data for this correspondence is collected not at public Internet services, but in internal IT infrastructure of businesses. Generalized analysis of received messages enables identifying the main problems in work with clients and carrying out multivariate analysis of clients in terms of their degree of satisfaction within messages or methods of interaction.

If a small business is looking for original value and significance of obtained results through web content processing, it will get a competitive advantage. Digital data becomes an economic resource if economic value is created on its basis, namely, collection, extraction, and processing of the data serve as a source for obtaining new value and creating a product or service. Nevertheless, a number of tasks may be distinguished that can be resolved on the basis of analysis of web content of social networks and services for opinion publishing:

(1) Study of public opinion, that is, on this basis, identification of check points in assessment of goods or services by consumers, determination of level of user satisfaction with the service, etc.

(2) Building of a taxonomy corresponding to the ideas of consumers and reflecting their conceptual picture, which helps to select an effective presentation of information, its categorization for the users' convenience.

(3) Monitoring of information field of the market to identify timely the phenomena and significant events that excite and disturb regional, expert, or consumer community.

METHODOLOGICAL APPROACHES TO AUTOMATED PROCESSING OF WEB CONTENT

Significant part of the web content is text in natural language. Large high-tech companies, as well as companies focused on work with users (Yandex, Facebook) and servicing the corporate sector (IBM, Oracle), are working for resolving the problem of automatic processing of the natural language. Machine learning technologies of neural networks and other data mining technologies have made a large contribution to resolving the problem of automatic processing of unstructured text [27]. An example of achievements in this area is Yandex or Google automatic online translating tools [28]. Generally, technologies of automatic processing of natural language are based on artificial intelligence systems and are inaccessible to a wide range of specialists who are not trained in the technical area and have no special hardware and software.

The service for automatic text analysis is present in information and analytical systems (IASs) for monitoring of media landscape in such systems as Medialogia, Integrum, SCAN Interfax, and Public.ru. IASs enable carrying out analysis of the content collected in their data warehouse, and information from social networks and blogs are represented in them selectively. Tools for text analysis embedded in IASs cannot be applied to arbitrary array of web content, and also do not contain setup mechanisms taking into account specific users' tasks. In addition, companies developing the IAS do not open the algorithms and methods of the text analysis used by them, since they were obtained as a result of complex setting up the machine learning of neural networks and represent a trade

secret. Due to high cost, consumers of IAS are large commercial and state organizations.

There exist specialized software package for text analysis, keyword identification, and automatic summarization[2]. However, these packages are not available to a wide range of specialists due to the need for special skills and usually used in large companies. Special applications for monitoring and processing of unstructured textual data of web content are marketed[3].

In educational materials which reveal technologies of automatic processing of texts in natural language, tasks are highlighted for a team of specialists which includes a data researcher, programmer, and subject matter expert, to competencies of which strict requirements are set. For example, a data scientist shall, on the one hand, do typical academic work, and on the other hand, participate in creation of a commercial product. The authors of [29] make the conclusion about immaturity of the applied side of data science on the basis of the explorative nature of most of the publications on this subject.

The modern level of IT enables to carry out efficient automatic processing of Big Data, including web content, but it requires involvement of high grade data specialists, programmers, as well as use of complex software tools or expensive services of developer companies. However, there are a number of text processing software tools available in public domain (open source) that can be used for resolving the tasks of small business. For instance, MyStem is a software [30] that can carry out morphological analysis of the texts in Russian and there is the Natasha library for processing the Russian texts in the Python software environment[4]. The software specified can be used at personal computers and, unlike commercial IASs, can form data arrays in accordance with individual business goals.

REQUIREMENTS OF SMALL AND MEDIUM BUSINESS FOR METHODS OF WORK WITH WEB CONTENT

Study and processing of web content open up new sources for creation of competitive advantages at the market, this is the reason for attention of large corporations to the Big Data. However, small businesses also need to be able to use this new resource in their business activity. Social media and digital platforms, on the one hand, serve as a tool for effective interaction of businesses and customers, and, on the other hand, as a source of valuable information. At the same time, processing of web content is complicated by difference in formats of unstructured messages and great number of data sources. To enable small businesses to use the new digital resource, web content processing tools shall correspond to their capabilities. Small business as compared to large one has limited resources, which prevents it from hiring high grade data specialists or purchasing of expensive software tools.

Official statistics data indicates that more than 90% of Russian businesses use personal computers, Internet access, and e-mail in their activities, and more than 80% use special software [31]. Level of small businesses informatization can be considered sufficient to carry out automated processing of web content using the tools that meet their computing, software, and personnel capabilities.

Methodological support of work with web content shall include automated tools, because in the conditions of increasing rates of accumulation of web content, labor costs for work therewith will also rise. It is required to adjust the work with web content in such a way that increasing amount of data would lead to an increase in the efficiency of their analysis, but not in the cost of their implementation. By now, the global market of information and analytical systems for work with the Big Data has been formed, but the offered facilities are not available for small businesses. The automation tools available for small businesses are the software distributed under a free license or embedded in popular office software packs.

In development of methodological approach to use of web content, it is necessary to take into account possibilities of involvement of human resources to the small business. The average level of monthly wages in small business (27569 rubles) is almost by 60% lower than the average level of wages in Russia (47867 rubles) in 2019 in general [32]. In most of cases, engagement of a data specialist by small business is impossible due to high cost of his/her labor at market: average compensation of such specialist according to assessment of recruiting site exceeds 100000 rubles per month[5]. Thus, the method for web content analyzing must be understandable and convenient for its effective use by any employee who has no special skills in work with data.

Practice of digital transformation of business shows that the key to its success lies not so much in possession of new technologies, but rather in development of economic mechanisms of extracting value from digital resources [1]. Essentially, the specialists employed in small business have unique knowledge to set task in work with data resolving of which will ensure the meaningful result.

Small business can benefit from an approach to work with web content that meets three requirements:

(1) use of free and affordable software,

[2] Tomita-parser (Yandex), ABBYY FlexiCapture, SDK Pullenti (Semantik LLC), IBM SPSS Modeler, EurekaEngine (PalitrumLab LLC).

[3] Kribrum JSC. https://www.kribrum.ru/. Data Analysis Laboratory by Aleksandr Kukushkin LLC. https://lab.alexkuk.ru/.

[4] Natasha library. https://natasha.github.io/ner/.

[5] Superjob. https://www.superjob.ru/z/.

Table 1. Steps of web content analysis

Step	Activity	Methodological and technological support	Result
The first	Extraction of web content data	Personal computer connected to the Internet, registration in web services	Sets of "raw" data
The second	Preparing (presentation)	Spreadsheets programs, data cleaning software	Database
The third	Analysis and interpretation	Office software suite, analytical platforms, and internet services	Analytical documents (new information in the form of tables, diagrams, and analytical reports)
The fourth	Application	Manager competencies	Managerial decision

(2) work using computational capabilities of a personal computer,

(3) no special training in the area of programming and data analysis for the specialists using this method.

Thus, we have developed a method of processing of web content for small businesses.

METHOD OF USE OF WEB CONTENT FOR SMALL BUSINESS

Taking into account requirements of small business, the method of use of web content consists of the activities described in Table 1 taken in four steps.

The purpose of the first step of web content analysis is in obtaining the data sets from selected sources of web content. It is important to determine clear criteria for assessment of the sources needed to resolve a management problem. For example, to make a decision on change of a service or a product, it is required to determine their characteristics and parameters that attract attention of the consumers. Accordingly, the research shall include the sources containing consumer opinions about a product or service.

High rate of web content accumulation makes it impossible to process the entire amount of data when resources of a business are limited. Therefore, the most valuable sources shall be selected and the information shall be extracted, processing of which will enable to fulfil the task formulated. An attempt of collection and analyzing of all possible data leads to spattering of limited resources, but not to increase in efficiency [33]. Sampling studies are the most preferable under the conditions of limited resources of small business. As compared to full study, a sampling study enables achieving high accuracy of results at lower costs. In addition, to check accuracy of sampling studies, statistical methods are applied, e.g. automatic functions of the widely used Excel.

Web services differ by terms for provision of access to web content. In some of them, it is required to send a motivated request for data extraction (Facebook), and in others it is required to effect payment for access (SCAN). A general requirement for work with data of all web services is registration as a user. Sources of web content are included in the study selectively, and data from them is retrieved using API [34], or by copying of content from a web browser, or using software for copying of a web site or part thereof (HTTrack Website Copier, etc.), so called parsing [35].

Results of the first step are raw data sets from each web service. The resulting datasets can differ by structure and presentation format depending on the source. However, the data structure in many sources includes the following fields: date, author, text, likes, and datasets are converted into a single format using data convectors, such as Power Query editor in Excel. Merging of data sets into a single format is required for subsequent automatic analysis thereof.

Tools for automatic text analysis provide processing of large amounts of data prepared in advance. Excessiveness of the natural language, a large number of word forms make automatic analysis of an unprepared body of text inexpedient [16]. For example, construction of frequency tables on the basis of words of text in the natural language leads to the fact that the same words are present several times in various forms, and give an erroneous idea of frequency of use thereof. An illustration of such an error is tag cloud in the area of quantum technologies (Fig. 1) in Atlas of End-to-End Technologies of the Digital Economy of Russia [36], which contains adjective "quantum" in all declensions in singular and plural.

At the second step of web content analysis, data is prepared in the form of a database suitable for analytical processing, including the processing with software tools. To do this, all words of the text shall be reduced to their original form, i.e., the lemma. For example, for the words "beautiful", "beautifully", or "beauties" will be found their common lemma "beauty". Counting of use of lemmas instead of words gives a more accurate understanding of proportion of the messages that mention such a characteristic of the product as

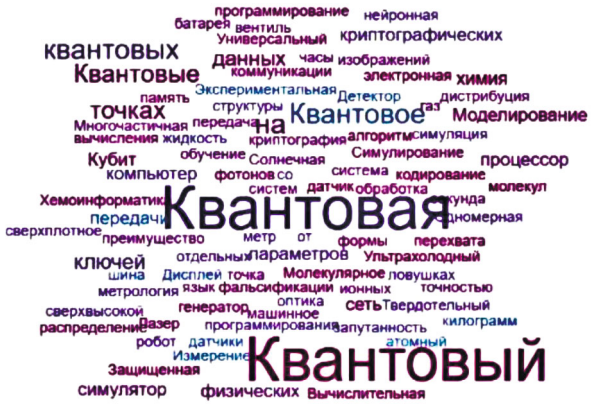

Fig. 1. Example of visual analysis of unprocessed textual data in Russian [36]. Mass Media (SCAN) Users' reviews.

"beauty." The MyStem[6] will fit for automated lemmatization of a text, it writes the result to a separate file of certain format, which is imported into the database.

The obtained data sets are entered into a common database in the tabular form, which will allow applying a widespread arsenal of analytical processing methods to them, including office software suite and special apps. Preparing of a set of textual data also includes removing neutral words which do not reflect the essence of the content, but are used often, the so-called "stop words". Stop words include prepositions, conjunctions, particles, pronouns, measurements units, and other words included in the stop word list[7] of the Russian, which, if required, may be supplemented with other words frequently occurring in the dataset, for instance, "such" or "nothing." For cleaning from stop words, the functions of filtering, sorting, searching, and replacing in spreadsheet editors are applicable.

As a result of execution of the second stage, the data extracted from the web services is transferred into a database form suitable for automatic analysis thereof, which enables answering the questions posed to the web content study to fulfil management tasks.

The third step of web content analysis includes analytical processing of data by software tools, formulation of conclusions, and recommendations for making managerial decisions. The studied web content is represented by simple texts, similar by style to colloquial speech, created by users (consumers, employees, managers) without the purpose of artistic expression, formulation of principles of law, etc. Simple analytical methods are used for the analysis thereof.

The statistical methods of analysis often used to resolve management issues are available to a wide range of users. Moreover, statistical analysis may be performed in popular spreadsheet. The statistical methods based on numerical and quantitative analysis ensure accuracy and objectivity of conclusions. Statistical analysis of words by frequency of their use and positioning in the text is carried out in the context of publication sources or other parameters.

The obtained results may be supplemented with calculated weights that take into account the number of users likes or comments to users' posts. In essence, at the third step, web content is transformed into some new information that can serve as an economic resource that enables to create economic value.

The fourth step of web content analysis is study and interpretation of the received information that require understanding of the essence of economic activity, the market and specificity of small businesses. At this step, methods, and tools are selected by specialists individually in accordance with the tasks set in the course of web content processing. Content analysis is used in scientific researches, for example, in sociology, for processing of large text arrays, when quantitative count of words, phrases, or other units in texts is performed with subsequent notional processing thereof [37]. In this regard the obtained results are supplemented with information from other sources.

TESTING OF THE SUGGESTED METHOD

For assessment of compliance of the proposed method of web content processing with requirements of small business, it was tested on the basis of example of the Moscow state budgetary institution Small Business of Moscow. The purpose of experimental web content analysis is determination of socially relevant characteristics of activity of an organization under study.

For selection of web content sources, popular services for publishing of reviews and opinions of users in the Russian segment of the Internet were studied, and the Internet services Google, Yandex, and Facebook were selected as containing regular reviews of organizations activity. This set did not include the sources containing isolated reviews of an organization under study. To assess interest of the journalistic professional community, publications of electronic mass media extracted from the information and analytical system SCAN were included. The analysis includes such reviews and subject matters for the period from October 2019 to September 2020. The compiled list of sources reflects the objective opinion of Internet users, since the selected Internet services are not associated with the organization under study.

Because the number of reviews in each source did not exceed several dozens, the datasets were formed by copying of content from a web browser and written down in the form of spreadsheets in Excel. Table 2 contains information on the obtained datasets.

[6] MyStem. https://yandex.ru/dev/mystem/.
[7] Stop words list. https://www.artlebedev.ru/yandex/site/saved/stopword.html.

Table 2. Characteristic of datasets

Internet services	Web content sampling for study		
	number of entries	average number	
		likes per one entry	words in one entry
Yandex	64	3.4	11.2
Google	22	1.0	16.5
Facebook	17	0.3	7.6
Total on reviews	103	2.3	11.8
SCAN (mass media subject matters)	100		

Then, all the extracted (collected) texts were broken down into separate words using the Excel text to columns function and written into a spreadsheet, in which the number of columns is determined by the number of words in the text. As a result of processing by MyStem, the resulting tables were written into new files, where the initial form (lemma) was provided for each word. Files with lemmas were unified into a single database.

Initially, the database contained 1.802 entries of word lemmas in the form of spreadsheet. After removing stop words, 1.314 words (lemmas) have remained. In Table 3, characteristic of the obtained database is provided.

In Excel, on the basis of the obtained database, statistical analysis of frequency of words use in the context of sources and reaction of other users to publications was carried out. Results of statistical analysis are provided in Fig. 2.

As a result of statistical analysis of web content, it is possible to make conclusions on which characteristics are distinguished by users and journalists in the organization under study, and, therefore, to estimate indirectly social relevance thereof, which represents solution to the set task.

In the experiment, the following means were used: a personal computer connected to the Internet, personal accounts in services Yandex, Google, Facebook, and VK, the information analysis system SCAN, and the software Excel, MyStem, and the web application "Oblokoslov.rf". Paid access to information resources and technologies was not required. Labor content required for processing of the web content, including steps of extraction, conversion and analysis of data made about four man-hours. Thus, depending on the data being studied, one specialist with advanced computer and Internet use skills will need no more than one business day to complete all steps of the web content analysis.

Table 3. General characteristic of database

Indicator	Mass media posts (SCAN)	Review from Internet users
Total entries	100	103
Total words (after cleaning)	646	668
Word stock	327	351
Mean frequency of use of the words	2	1.9

In our study, the following restrictions of the proposed methodological approach were revealed. The following is not taken into account: words with typos and errors, probable publications for the purpose of unfair competition, the so-called fake engagements, and moderation of Internet services. Reviews and comments pass moderation for adherence to ethical rules and the dictionary used. The moderation that cuts off non-constructive posts from a publication facilitates analysis of the web content. However Internet services do not disclose moderation algorithms to complicate publication of automatic reviews content by robots. But in this case analysts cannot estimate which share of reviews was lost due to moderation.

The method of work with web content suggested by us demonstrated its applicability under conditions of limited access to information competencies and technologies and proved its effectiveness on the basis of the example considered, because it gave an answer to the test task formulated.

CONCLUSIONS

Digitalization of the society expressed in intensive distribution and use of information technologies forms new kinds of resources for creation of value in economic activities of businesses. One of such resources is users' web content, for processing of large volumes of which, large high-tech companies have initiated large-scale investment projects, while a significant part of the country economic entities, mostly small businesses, have limited opportunities to use achievements of digitalization to the full extent. The business which generates more than 20% of GDP of Russia, cannot compete with large participants by the

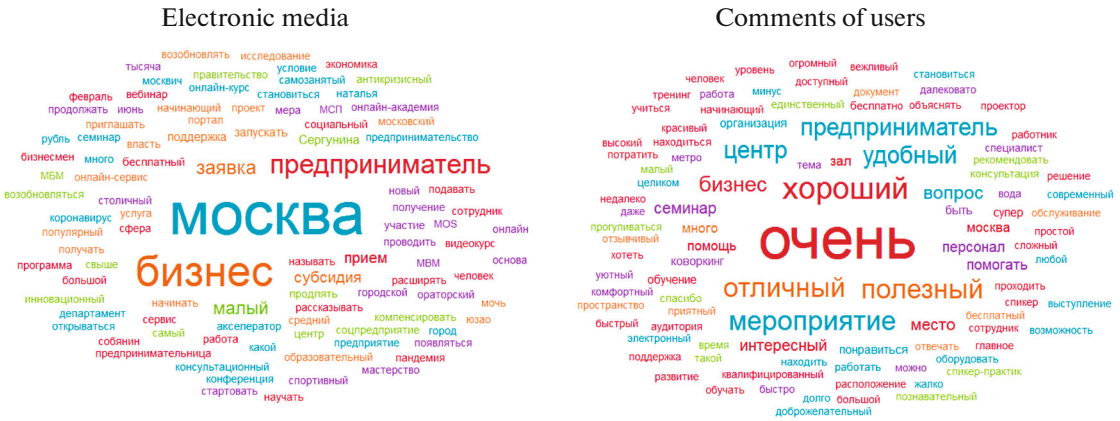

Fig. 2. Visual analysis of words used in sources (in Russian).

volume of web content being processed. And for the development of digital economy, it is important that new opportunities could be used by a wide range of economic entities. It shall ensure growth of the country economy.

For processing of web content as an economic resource suitable for use by small business, a methodological approach shall be developed. On the one hand, the small business cannot engage high-end professionals or acquire/create expensive IT infrastructure, and, on the other hand, it has unique knowledge in the subject matter required for extraction of value from the data obtained. Selective processing of web content using the technologies and methods available to small business will enable to resolve a number of tasks, such as finding out of new needs, clarifying of taxonomy or monitoring of the information space. Individual technologies of data processing become available for small business due to reduction of cost thereof and making qualification requirements less strict.

The method developed by us comprises sampling study of web sources and consists of four steps. At each step, relevant conclusions are made which are of independent importance for information issues resolving. The results of study of web information sources and the developed criteria for selection of information may be in demand for establishment of an automatic monitoring system. Technological support of the method includes spreadsheets and freeware, and, consequently, it does not require any extra financial expenses. The method contains data processing using embedded functions and formulas and does not require programming, i.e., it is understandable for specialists having basic information competencies. Nevertheless, the proposed approach is not complete to ensure obtaining of economic benefits by small business. It shall be supplemented by a data processing task that can be executed by specialists in the subject matter. Resolving of this task will ensure obtaining of valuable data.

The performed testing of the developed method which does not require high costs and labor content, demonstrated its critical importance for small businesses, as it enables use of benefits of the society digitalization. Further elaborations for development of mechanisms of work of a small business with web content may utilize machine learning or neural networks methods as the next step, subject to sufficient extraction and accumulation of digital data.

REFERENCES

1. Ananyin, V.I., Zimin, K.V., Lugachev, M.I., Gimranov, R.D., and Skriprin, K.G., Digital organization: Transformation into the new reality, *Bus. Inf.,* 2018, no. 2, pp. 45–54. https://doi.org/10.17323/1998-0663.2018.2.45.54
2. Moazed, A. and Johnson, N.L., *Modern Monopolies, What It Takes to Dominate the 21st Century Economy,* New York: St. Martin's Press, 2016.
3. Tapscott, D., Macrowikinomics, *Rebuting Business and the World,* Penguin, 2012.
4. Lugachev, M.I. and Skripkin, K.G., Information revolution: The economic aspect, *Vestn. Mosk. Univ., Ser. 6, Ekon.,* 2019, no. 6, pp. 20–38.
5. Abdrakhmanova, G.I., Vishnevskii, K.O., Gokhberg L.M., et al., *Indikatory tsifrovoi ekonomiki: 2020. Statisticheskii sbornik* (Indicators of the Digital Economy: 2020. Statistical Yearbook), Moscow: Nats. Issled. Univ. Vyssh. Shk. Ekon., 2020, p. 196. https://www.hse.ru/primarydata/ice2020.
6. Mar, B. and Ward, M., *Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems,* United Kingdom: John Wiley & Sons, 2019.
7. Mytenkov, S.S. and Markova, E.S., In pursuit of the digital future: Analysis of the effectiveness of small and medium-sized businesses in the field of ICT for the implementation of the program "Digital Economy of the

Russian Federation," *Biz. O-vo. Vlast',* 2018, no. 2, pp. 159–174.

8. Nissen, V., Lezina, T., and Saltan, A., The role of IT-management in the digital transformation of Russian companies, *Foresight STI Gov.,* 2018, vol. 12, no. 3, pp. 53–61. https://doi.org/10.17323/2500-2597.2018.3.53.61
9. Stolypin Institute of Economic Growth, Sector of Small and Medium-Sized Enterprises: Russia and the World, 2018. https://stolypin.institute/novosti/sektor-malogo-i-srednego-predprinimatelstva-rossiya-i-mir/. Cited January 18, 2021.
10. Li, Y. and Rama, M., *Firm Dynamics, Productivity Growth, and Job Creation in Developing Countries: The Role of Micro- and Small Enterprises,* Oxford Univ. Press on behalf of the World Bank, 2015. https://openknowledge.worldbank.org/handle/10986/24807. Cited January 18, 2021.
11. Filimonova, N.M., Morgunova, N.V., and Sinyavskii, D.A., Determination of promising directions for the study of small and medium-sized businesses, *Nauchno-Tekh. Inf., Ser. 1,* 2014, no. 9, pp. 20–26.
12. Smirnykh, L.I., Digital literacy of the elderly and digitalization of enterprises: Experience of European countries, *Vopr. Ekon.,* 2020, no. 12, pp. 104–124.
13. Syuntyurenko, O.V., The risks of the digital economy: Information aspects, *Sci. Tech. Inf. Process.,* 2020, vol. 47, no. 2, pp. 104–112. https://doi.org/10.3103/S0147688220020082
14. World Bank Group, *World Development Report 2016: Digital Dividends,* Washington, DC: World Bank, 2016. https://openknowledge.worldbank.org/handle/10986/23347. Accessed January 18, 2021.
15. Gavrilova, T.A., Kudryavtsev, D.V., and Kuznetsova, A.V., The choice of knowledge management tools taking into account the specifics of the subject area, *Innovatsii,* 2019, no. 8, pp. 44–52. https://doi.org/10.26310/2071-3010.2019.250.8.007
16. Belonogov, G.G., Gilyaresvkii, R.S., Seletkov, S.N., and Khoroshilov, A.A., Ways to improve the quality of textual data searches on the internet, *Autom. Doc. Math. Linguist.,* 2013, vol. 47, no. 4, pp. 111–120.
17. Syuntyurenko, O.V., Theoretical and applied aspects of automating multivariate analysis procedures, *Autom. Doc. Math. Linguist.,* 2018, vol. 52, no. 6, pp. 275–281.
18. Eremenko, K., *Rabota s dannymi v lyuboi sfere: Kak vyiti na novyi uroven', ispol'zuya analitiku* (Working with Data in Any Field: How to Reach a New Level Using Analytics), Moscow: Al'pina Pablisher, 2019.
19. Batura, T.V. and Bakieva, A.M., *Metody i sistemy avtomaticheskogo referirovaniya tekstov* (Methods and Systems for Automatic Text Summarization), Novosibirsk: IPTs NGU, 2019. https://www.iis.nsk.su/files/book/file/Batura_Bakiyeva_Autosummarization.pdf.
20. Tarasov, S.D., Modern methods of automatic abstracting, *Nauchno-Tekh. Vedomosti S.-Peterb. Gos. Politekh. Univ., Inf. Telekommun. Upr.,* 2010, no. 6, pp. 59–74.
21. Batura, T.V., Methods for automatic classification of texts, *Program. Prod. Sist.,* 2017, vol. 30, no. 1, pp. 85–99.
22. Efremova, M.I., Automatic parsing and annotation of articles, *Fundam. Issled.,* 2015, no. 2, pp. 4866–4870.
23. Dneprovskaya, N.V., Investigation of the transition of an enterprise to the digital economy, *Vestn. Ross. Ekon. Univ. im. G. V. Plekhanova,* 2019, no. 4, pp. 54–65.
24. Kalabikhina, I.E. and Chesnokov, A.A., Using content analysis of communications in an online consultant to improve customer experience, *Gos. Upr., Elektron. Vestn.,* 2020, no. 80, pp. 151–174.
25. Yatsko, V.A., A system for automatic analysis of customer opinions, *Prikl. Inf.,* 2015, no. 4, pp. 5–17.
26. Tverdokhlebova, M.D., Skorobogatykh, I.I., Musatova, Zh.B., and Nevostruev, P.Yu., Assessment of consumer preferences with the use of user content analysis using the example of electronic book products, *Mark. Issled.,* 2020, no. 3, pp. 190–202.
27. Mosyagin, A.B., Using the data mining methodology in solving the problems of processing social data, *Monit. Obshch. Mneniya: Ekon. Sots. Peremeny,* 2015, no. 3, pp. 143–146.
28. de Vries, E., Schoonvelde, M., and Schumacher, G., No longer lost in translation: Evidence that Google Translate works for comparative bag-of-words text applications, *Polit. Anal.,* 2018, vol. 26, no. 4, pp. 417–430.
29. Bengfort, B., Bilbro, R., and Ojeda, T., *Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning,* O'Reilly Media, Inc., 2018.
30. Segalovich, I., A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine, *Proceedings of the International Conference on Machine Learning; Models, Technologies and Applications (June 23–26, 2003),* Las Vegas, pp. 273–280.
31. Federal State Statistics Service. Information Society. Information and Communication Technologies. https://rosstat.gov.ru/folder/14478. Cited January 18, 2021.
32. Federal State Statistics Service. Labor Market, Employment, and Wages. https://rosstat.gov.ru/labor_market_employment_salaries. Cited January 18, 2021.
33. Veduta, E.N. and Dzhakubova, T.N., Big Data and economic cybernetics, *Gos. Upr., Elektron. Vestn.,* 2017, no. 63, pp. 43–66.
34. Shevtsova, I.V., Methods for teaching to work with digital data, *Otkrytoe Obraz.,* 2020, vol. 24, no. 4, pp. 32–40.
35. Shvedov, D.I., Parsing and data matching for operational analysis of dynamic situations, *Gorn. Inf.-Anal. Byull.,* 2017, no. 6, pp. 200–204.
36. Makushin, A.G. and Osochenko, E.A., *Atlas skvoznykh tekhnologii tsifrovoi ekonomiki Rossii* (Atlas of End-to-End Technologies of the Digital Economy of Russia), Moscow: AO Grinatom, 2019.
37. Trotsuk, I., When methodology beats techniques; or, why we prefer discourse and narrative analysis to interpret textual data, *Russ. Soc. Rev.,* 2015, vol. 14, pp. 48–63.