The Description of the Structure of Social Identity in the Information Space, Using Automated Data Processing Tools

Boris Nizomutdinov[1][0000-0002-4090-9564], Anna Uglova[2] [0000-0002-8072-0539], Aleksandr Tropnikov[3] [0000-0002-4179-536X]

[1,3] ITMO University, Kronverksky Pr. 49, 197101, Saint-Petersburg, Russia
[2] Herzen University, 48 Moyka Embankment, 191186 Saint-Petersburg, Russia

boris-wels@yandex.ru

Abstract. Recently, the topic of analyzing digital social identity has become more relevant all over the world. The article presents the results of the second stage of a research project involving the development of a methodology for automated analysis of digital social identity based on the VKontakte social network in order to study the relationship between the visual component of the social profile and the psychological characteristics of respondents. To achieve this goal, a comparative analysis of the tools based on the use of machine learning technologies for the automated analysis of text and visual data that form the basis of the information image of social identity was carried out. Using cluster analysis, visual identity strategies were identified. To identify digital factors mediating the formation of social identity, a correlation analysis of data obtained by means of automated analysis of graphic data and the results of psychodiagnostic research was used. Visual identity strategies have formed multiple relationships with the psychological characteristics of users. Based on the obtained relationships, a conclusion was made about the broad possibilities of automated analysis of digital social identity.

Keywords: machine learning, cluster analysis, graphical data analysis, social networks, social identity, visual self-presentation.

1 Introduction

At the present stage of development of social and psychological sciences, there is a lack of scientific research devoted to the analysis of the formation of social identity in the changing socio-cultural environment and the digitalization of modern society. Immersion in the Internet space has a strong influence on the formation of motivational, regulatory, and reflexive spheres of the user's personality and can become both a protective factor and risk factors for the destruction of social identity [5, 17]. The search is relevant due to strengthening the impact of virtual images on real life and the network nature of

Copyright © 2020 for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR-WS.org/Vol-2784/rpaper18.pdf

almost all social interactions of modern people, the automated analysis of which will allow analyzing the social identity of users in real time [2, 12].

We can say that modern socio-psychological research needs to update the methodological model of experimental study of social identity through an interdisciplinary approach and the introduction of modern information technologies in the process of analyzing the forms and mechanisms of personality presentation in the virtual space [10]. Social networks are the most promising platform for exploring the inner world of users, as well as individual and group social interactions [7]. The majority of Russian studies that study people in the information space rely on traditional psychodiagnostic research methods. This research is aimed at combining traditional psychodiagnostic methods and modern automated methods of data processing and collection to further predict the socio-psychological characteristics of the user based on the analysis of their social profile without using time-consuming psychodiagnostic techniques [11].

Modern social networking sites are databases of publicly available information that users themselves fill in. This information can provide a basis for analyzing socio-psychological characteristics, identifying vulnerable social groups (people with suicidal tendencies, people with antisocial habits and behavior) [1, 4].

Modern digital information processing technologies can be used to facilitate the identification of these and many other human characteristics. Automated data collection, parsing of personal pages, and machine image processing-all these technologies allow you to speed up and expand the methods of analyzing users of social networks [14, 15].

In our pilot study, where the information image was analyzed, a theoretical model of automated data processing of social network users was developed [6].

The purpose of this work is to develop a set of automated tools for analyzing the factors of digital social identity formation, based on the analysis of text and visual information, and to correlate the data obtained in the automated mode from the social profile with the results of a psychodiagnostic experimental study of subjects to identify significant relationships between their personal traits and elements of the information image.

2 Research methodology

Currently, there are practically no interdisciplinary Russian socio-psychological studies devoted to automated analysis of the structure of digital changes in the social identity of an individual. Existing international research provides a methodological basis for combining psychological, sociological and informational approaches that will describe the digital environment as one of the basic areas of life that set the direction of modern human development (Mossberger K., Soldatova G.) and answer the question of the degree of influence information technologies on the psychological state of the user, his interpersonal, group and social relations [17].

2.1 Methodological model of the study

The study was conducted in several stages:

- The task of the first stage was to analyze tools and programs designed to collect and analyze data from their social networks. A comparative analysis of functions was required to select suitable software and services options.
- At the second stage, subjects were selected to conduct a survey and an experimental study of socio-psychological characteristics. A psychodiagnostic study was conducted. In addition, the respondents gave their consent to the analysis and processing of their profile in the social network Vkontakte.
- At the third stage, the authors parsed the profiles of users in the Vkontakte social network who underwent a psychodiagnostic study at the second stage. Special services were used for uploading, as well as available social network APIs. All processing was carried out with the consent of the respondents and did not violate the user's personal data processing policy. In the course of discharge were screened inactive or disabled users, it is possible to reduce the inaccuracy of the data. This stage allowed us to select the appropriate tools for data collection. In future research, we plan to develop a script that allows you to automatically exclude unsuitable profiles for analysis.
- At the final stage, using the methods of mathematical statistics, using cluster and correlation analysis, the relationships between the uploaded data from social profiles and the socio-psychological characteristics of users were revealed.

The study involved 176 people belonging to the first period of adulthood (19–32 years). The average age of the subjects is 25.3 years. The study involved 58 % (103) women and 42 % (73) men.

The traditional psychodiagnostic method was used to study the psychological characteristics of the subjects, mediating the formation of social identity. We have developed a set of psychodiagnostic tools:

- Maslow's needs satisfaction diagnostic test, which revealed the hierarchy of an individual's basic needs that underlie social identity;
- C. Schwartz's value questionnaire, aimed at analyzing social and individual values that determine the direction of interaction in the network;
- Self-presentation tactics scale (Lee S.-J., Quigley B., Nesler M., Corbett A., Tedeschi J.), which defines the leading strategies and tactics of self-presentation in the information space.

The following automated methods were used in the study of a user's profile in a social network:

- methods of automated collection of publicly available information from profile pages in the Vkontakte social network. As a result of data collection, an array was obtained that represented statistical information about the user's social network profile. Two approaches were used to collect information from social networks:

  - Parsing basic information of user profile pages in a social network vk.com, based on collecting information via the API, as well as uploading HTML content, followed by parsing and searching for the necessary profile information. At the moment, the privacy policy allows you to hide the main information, but even if all information is hidden, you can upload your avatar, name, number of subscriptions, songs, photos and videos. If the profile is open, the list of available information is much wider.
  - classification Methods based on the supervised learning approach and semi-supervised learning approach. The essence of the task is to correlate a set of objects/features with a predefined set of classes/categories. In this study, they were used to automatically classify communities/groups in the Vkontakte social network that a user belongs to by various topics in more than 30 areas (Science, Design, Humor, Entertainment, Games, Travel, etc.);
- methods for processing graphic and text data for automated analysis of graphic data, aimed at identifying a number of specified parameters: the number of people in the photo, people's emotions, objects present, composition, color correction, blurring, etc.

2.2 Methodological analysis of the tools

There is a certain array of developments based on the use of machine learning technologies in terms of the tools for the proposed research. Machine learning is a broad section of the field of artificial intelligence, that studies methods of building learning algorithms. There are three broad approaches in machine learning depending on the type of feedback signals or data, transmitted to the training system [16]:

- supervised learning and semi-automated learning: the machine receives a sample of input data and expected results, set by the "teacher", the aim of the system is - to study the general rules of bringing input data to the final result;
- unsupervised learning: the system tries to find some structure in the data pattern by itself and then to make some conclusion, and this approach does not provide for premarked data sets.
- reinforcement learning: a software application interacts with a changing environment, in which the system must perform some specific task without preliminary indicating the final result.

At present time machine learning methods solve a fairly wide range of the following standard tasks:

- classification task (Sokolova M., Lapalme G.) is the most popular type of tasks in machine learning, it is a mapping of a certain set of objects with class labels based on a given finite training sample, divided into classes beforehand. There are binary and multi-class classifications, as well as disjoint, intersecting, and fuzzy classes;
- regression task differs from the classification, that the end result is a real number or a numeric vector (a gap);

- learning to rank task has a special feature in getting a lot of output data and automatically sorting by response values; it is often used in information search and text mining;
- forecasting task is a set of objects that are time series segments that end at the moment when it is needed to make a forecast for the future value;
- clustering task (Bakker B., Heskes T.) is to divide objects into clusters (disjoint groups), using data on the similarity of paired elements and common signs;
- outliers detection task is to detect a small number of objects that deviate from the norm of the training sample; this approach allows to get rid of unwanted noises in the sample, inaccuracies or errors in the data;
- missing values task is to replace the missing values in the object matrix with their predicted values.

Machine learning as an interdisciplinary field includes such related areas of knowledge, as mathematical statistics, optimization methods, information extraction methods, data mining. Research in the field of machine learning is carried out by conducting experiments on model or real data to check the correctness of methods, confirm hypotheses, get a list of statistically significant criteria, calculate statistical metrics.

As central methods (which are also often called approaches or algorithms) of machine learning are the following (Zhang C., Ma Y., Dietterich T.G.): linear and logistic regression; SVM (support vector machines); decision trees; random forest; Naive Bayes; boosting; neural networks; deep learning; K-means; KNN (k-nearest neighbors); self-organizing maps etc.

Each of the presented methods has its own advantages and disadvantages, therefore they can be used to solve completely different types of problems. Natural Language Processing (NLP) it is used in combination with machine learning methods to perform tasks such as, emotion detection, text tonality analysis, speech recognition, spam classification in emails, machine translation, speech recognition, and so on. NLP plays a very important role in collection, processing and analysis data by converting natural language into a format, that further used by machine learning methods to implement its own algorithms. Thus, Goldberg Y. in his works, and also Batura T.V. in a review study of automatic text classification methods and Zibert A.O. with Hrustalev V.I. provide the main methods and approaches of natural language processing (NLP) in the frames of working with neural networks of different architectures and with standard statistical models for implementing deep learning methods, they also present the main results, obtained during the implementation of these methods.

Among such methods and approaches can be distinguished the following: tokenization; making a list of stop words; stemming; lemmatization; Named Entity Recognition; “bag of words” model; function calculation TF-IDF; Word2Vec algorithms; etc.

Based on a number of ongoing studies in the field of identifying the relationships of social networks and personal characteristics of users (Kosinski M., Zhang C., Settanni M., Azucar D., Marengo D.), we can identify several areas for determining the relationship of data obtained from social networks and personal qualities of users [9]:

- processing of photo images
- semantic analysis of text user "posts"

- analysis of statistical data.

Collecting information of this scale, as well as analyzing it, requires considerable effort. One of the ways to speed up this process is to use cloud computing technologies that can remotely accept and process requests for analysis of various data: both text and graphic. They can be used to perform both complex cognitive transformations on data (search for individual objects in the image, compose text descriptions of photos), and complex analytical functions (counting the frequency of word use, highlighting the prevailing speech patterns, determining the subject of messages).

Such cloud platforms are developed by large companies that specialize in software development and distribution: Microsoft, Google, and Amazon. Their cloud solutions provide various functions that can automate the process of analyzing a user's profile in social networks to varying degrees.

Microsoft's multi-functional Azure cloud platform is one of the largest open services offering remote data processing services. The functionality of the service is designed for individual users who use the capabilities of computers for personal purposes, as well as for entire companies that conduct extensive monitoring studies.

The core module of Azure is the intelligent cognitive Services interface, which combines most of the complex and complex functions of machine data processing. Among them: face recognition (for the presence of certain emotions, estimated age, skin color), computer vision (analysis of photos for the presence of specific objects, drawing up text descriptions) [3].

Over the past 4 years, several large and dozens of pilot studies have been conducted related to the application of the functionality of this platform. Most of these studies were aimed at identifying various relationships between the identified characteristics of photos and personal traits of a person. For example, various correlations were established between the technical data of the image (blurring, the presence of a face in focus, the photo's sharpness) and the age of the person, the relationship between the number of faces in the photos and the personality traits of extroverts [8].

Working with Microsoft Azure Cognitive Services was performed using the API connection using the available methods. A special script was prepared for this purpose. The service allows you to perform 20 transactions per minute, with a total of 5,000 and 50,000 transactions per month for the Computer vision and face Recognition services, respectively.

Google also provides services for the allocation of computing power and services United by a common Google Cloud Plat-form platform. As a direct competitor to Microsoft and their Azure platform, Google provides similar data processing services: the Cloud product Cloud Vision repeats most of the functions of Azure and is used for complex operations on graphic and text data (image analysis, pattern detection, machine text analysis).

Like its counterparts, Cloud Vision is able to identify faces in images, analyze the emotions displayed, capture distinctive features in the form of hair or headwear, identify objects and situations in the background, and create a set of "tags" for images. Unlike Microsoft Azure, Google Cloud Vision specializes in analyzing existing objects when processing an image, rather than simply fixing them. Thus, the set of final tags

for the two services differs: the Microsoft service lists all possible objects in the image and displays a brief description of what is happening, while the Google service displays an already analyzed set of tags obtained from certain objects. Amazon Web Services is a platform for providing machine data processing services from Amazon. This platform specializes in providing services for organizing cloud databases, serverless computing, development tools, virtual servers, and storage. Currently, Amazon Web Services is focused on providing infrastructure and platform services. However, AWS provides Amazon Rekognition technology for image and video mining.

Amazon Rekognition provides the standard functionality of the machine image processing service. You can use it to determine faces, objects, tags, the degree of decency of the image, the color scheme, and a number of other technical characteristics.

In the course of conducting experimental testing and studying analytical articles of consulting companies, it was found that this service is inferior to the described two previous ones in the accuracy of determining the displayed objects.

Amazon Rekognition can only capture objects in an image, but not analyze the overall picture, unlike Google Cloud Vision. Also, the technology from Amazon is not able to make a meaningful description of the image as a competitive service from Microsoft. However, the overall accuracy of fixed tags is relatively low.

Based on a number of studies (Sophie W.F., Xenos S., Ryan T.) [9, 13, 18], it can be established that linguistic features can be used to recognize personal characteristics. The described methods can be used not only for analyzing handwritten or typewritten texts, but also texts left by users of social networks.

Most of the research conducted on this topic uses the program “Linguistic Inquiry and Word Count", which allows you to calculate the proportion of certain parts of speech in the text, the number of words longer than a certain value, the number of punctuation marks used, and the number of words from various lexical and semantic categories. Further, the data obtained are correlated with the data obtained during testing of the authors of the text and during processing, the correlation between certain linguistic features and personal characteristics is highlighted.

The main problem is the lack of publicly available and well-developed Russian language libraries for this type of program. However, using online services that are similar in functionality, you can process more text information with less effort.

Microsoft Azure, previously described as a platform for providing cloud computing services, also has a text message analysis service. The data obtained can be used to determine a person's attitude to certain objects or events, the subject of their text messages, the range of interests and the General tone of publications.

ISPRAS API is a non-commercial product developed by the Ivannikov Institute of system programming of the Russian Academy of Sciences. This product is specialized in natural language processing and analysis. Currently, there are several demos of various software solutions, one of which is aimed at semantic text analysis – Text Processing. This software product is implemented on a non-commercial basis and offers more impressive functionality than its analogues. The data obtained during the processing of text messages can be used not only to determine the user's Hobbies or their relationship to certain objects, but also to find relationships between the frequency of use of certain parts of speech and personal traits.

One of the most well-known companies engaged in machine analysis of user data was Cambridge Analytica (CA), created on the basis of research by Kosinski M. Kosinski M. conducted research using social media apps, asking users to take various psychological tests. By collecting users ' data with their permission, he was able to prove that there is a link between a person's online activities and their real "alter ego". For example, using 68 "likes", you can determine the user's gender, age, sexual orientation, and political preferences with a certain degree of probability. Based on these studies, the system used by the CA was developed. Using user data from social networks and based on the results of a study by Kosinski M., SA specialists were able to predict their information images. Further, information images were segmented according to various parameters (political preferences, psychological traits, etc.), after which data about sorted users were sold to third companies to demonstrate targeted political advertising that was created specifically for certain categories of people.

Using this experience, many Russian companies began to offer their services for identifying and segmenting the user population. So, the head of Sberbank said that the company intends to use the methods of Kosinski M. to identify customers with a high risk of late payments.

At the moment, there is a wide range of solutions on the Russian and international market, which is represented by various companies, such as: Palantir Technologies, Cambridge Analytica (abolished on 01.05.2018), Me-dialogia, Search-IT, I-tech, etc.

The solutions provided by these companies are used in various business areas

- credit scoring in banking organizations;
- recruitment of personnel in HR departments;
- personalized advertising;
- various types of recommendation systems.

In addition to commercial applications, these technologies are also used in political agitation, propaganda, and identification of potentially dangerous or vulnerable social groups

3 The results of the study

At the first stage of automated analysis of social network profiles the visual content of users using the Microsoft Azure Cognitive Services platform was analyzed, which allowed to identify 227 tags, found on users' avatars. We selected the main 46 tags (objects), that are found in the majority of subjects. The tag data set represents the basic visual identity of Vkontakte users, which can become a unique source of information about a person's social life, gender, age, and political characteristics for various types of socio-psychological research.

To identify the main strategies for building visual identity, presented in photos that combine tags obtained during automatic analysis, cluster analysis (Ward's method) was used (Figure 1).

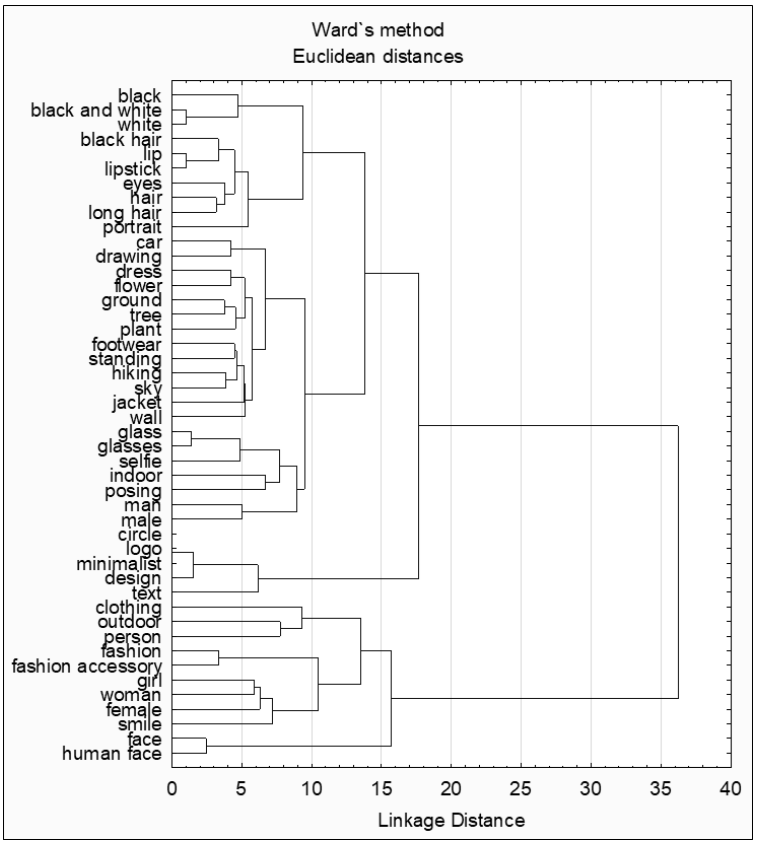

Fig. 1. Categorical structure of visual identity of users of the Vkontakte social network.

Cluster analysis allowed us to identify and describe the main strategies for building visual identity in the social network:

- Business, portrait self-presentation (black, black and white, black hair, white, lip, lipstick, eyes, long hair, portrait, hair), intended to create a certain distance between the user and the viewer;
- Visual storytelling (car, drawing, dress, flower, ground, plant, tree, footwear, standing, hiking, sky, jacket, wall), intended to convey an emotional and imaginative message through the natural surroundings, to please audience
- Staging of feminine self-presentation (clothing, face, fashion, fashion accessory, girl, human face, outdoor, person, smile, woman, female), it reflects a regulated socially-oriented identity and presents a set of popular clichés, to create a normative image;
- Staging of male self-presentation (glass, glasses, selfie, indoor, posing, man, male), it is also a reflection of a regulated socially-oriented identity;

- Anonymous self-presentation (design, logo, minimalistic, text), intended to indicate the user's unwillingness to expand the circle of social contacts, aimed at a narrow circle of users.

At the next stage using correlation analysis we identified reliably significant connections between the identified strategies for building visual identity and elements of real social identity studied in the course of psychodiagnostic research:

- Business, portrait self-presentation is more often used by subjects who spend a large amount of time on a social network (r=0,16, p⩽0,05), with a strong need for professional development and self-actualization (r=0,22, p⩽0,05);
- Users who use the "visual storytelling" strategy are more likely to indicate that their profile matches their real Self (r=0,17, p⩽0,05), the value of self-esteem is important to them (r=0,17, p⩽0,05), it is important for them to demonstrate the attributes of their identity in the photo. Also the value of unity with nature (r=0,18, p⩽0,05) is significant to them what is successfully demonstrated through visual self-presentation.

Users focused on staged female self-presentation, use a large number of social networks (r=0,19, p⩽0,05), to manifest the preferred values of “finding the meaning of life” (r=0,17, p⩽0,05), "values of self-respect» (r=0,26, p⩽0,05), “the values of unity with nature” (r=0,16, p⩽0,05), "the values of accepting life» (r=0,16, p⩽0,05), “the value of honesty” (r=0,19, p⩽0,05), “self-affirmation needs” (r=0,19, p⩽0,05), “the need for self-actualization” (r=0,17, p⩽0,05), “self-value” (r=0,19, p⩽0,05). We can say that this group of users actively develops the informational space, that merges with their real life and is used to assert their value and significance as a person.

Users, who use staged male self-presentation choose such behavior tactics as “bullying” in interaction with other people (r=0,16, p⩽0,05), they do not see value in the “sense of belonging to a group” (r=-0,17, p⩽0,05), for them other people opinion is not significant (r=-0.20, p⩽0.05), they do not feel the need to be modest (r=-0.22, p⩽0.05), which may indicate a desire to present a traditional stereotypical masculine identity.

Anonymous self-presentation is preferred by users who use a minimal number of social networks (r=-0.17, p⩽0.05), for whom the value of freedom is important (r=0.16, p⩽0.05) and the “value of self-respect” (r=-0.22, p⩽0.05), “value of obedience” (r=-0.19, p⩽0.05) and “self-acceptance” (r=-0.18, p⩽0.05) are not significant.

At the next stage, for a qualitative value-semantic analysis of the content of users' social identity, we performed an automatic classification of communities/groups in the Vkontakte social network that the user belongs to. There were 31 main categories that can be assigned to groups that are most frequently visited by users. Categories, in turn, can be divided into 9 groups of values, that form the basis for the formation of social identity:

1. Professional values (professional communities, job search);
2. Traditional family values (family, parenting, home improvement);
3. Value relationships (community for dating, finding a partner).
4. Values of social communication (local communities, communities of interests);

5. The value of diversity, of novelty (travel, leisure, entertainment, community events);
6. Information values (mass media, public pages);
7. Hedonistic values (entertainment news, watching movies online, photos, music groups, online stores, services);
8. Values of personal growth (scientific communities, creative communities, literary communities);
9. Values of beauty and health (information about healthy lifestyles, nutrition, sports, beauty).

After analyzing the percentage of communities in which network users most often belong, we can say that users most often belong to local communities that manifest their territorial affiliation to a social group in a particular district, city, or country (23.37% of the subjects). The second place is occupied by professional communities (17.39% of the subjects), which corresponds to the need of the subjects of the first period of mature age (20-30 years) to develop and consolidate their professional identity. The third place was taken by communities related to the satisfaction of hedonistic needs (groups dedicated to cooking 11.24%, online stores 10.97%), which indicates the close interweaving of their real and virtual lives and the use of social networks to organize their social space. The fourth place is shared by groups dedicated to creativity (9.71%), family (8.50%), and design (8.49%).

Thus, we can say that belonging to certain communities and visual self-presentation reflect the internal solidarity of the user with certain ideals of modern society and allows us to describe social identity as a multi-level structure that is reflected in the information space.

4 Conclusion

In general, the study of digital social identity corresponds to the current task of fundamental psychology, which is to develop new approaches to the study of patterns of personality formation in the information society. The existing range of needs of socio-psychological and social practice in science-based technologies for supporting digital communication relates to the need to create and test tools that allow you to build predictive models of behavior on the Internet that leads to the formation of social identity.

Research interest in the subsequent stages of the work should be focused on finding non-obvious patterns and correlations in the source data by using machine learning methods to speed up the getting and processing of information about social identity in the information space, as well as evaluating the predictive reliability of the model based on conducting and analyzing the results of a longitudinal study.

References

1. Cote, J.E., Schwartz, S.J.: Comparing psychological and sociological approaches to identity: Identity status, identity capital, and the individualization process. Journal of Adolescence, (25), 571–586 (2002), https://doi.org/10.1006/jado.2002.0511, last accessed 28.11.2020.

2. Emelin, V.A.: Vysshie psikhicheskie funktsii v kontekste tsifrovykh. Tsifrovoe obshchestvo v kulturno-istoricheskoi paradigme, kollektivnaia monografiia. Pod redaktsiei T.D. Martsinkovskoi, V.R. Orestovoi, O.V. Gavrichenko. Moskva, 177–181 (2019), https://istina.msu.ru/projects/8804361/, last accessed 28.11.2020.
3. Kaiqi, H., Qiao, W., Zhenyang, W.: Natural color image enhancement and evaluation algorithm based on human visual system. Computer Vision and Image Understanding, I (103), 52–63 (2006), https://doi.org/10.1016/j.cviu.2006.02.007, last accessed 28.11.2020.
4. Katz, J.E., Rice, R.E.: Social consequences of Internet use: Access, involvement and interaction. Cambridge, MA: The MIT Press, (2002), https://mitpress.mit.edu/books/social-consequences-internet-use, last accessed 28.11.2020.
5. Letov, E.V. Setevaia identichnost v kontekste kulturnykh protsessov informatsionnogo obshchestva: avtoreferat diss. … kand. filosof. nauk., 19 s. (2014), https://static.freereferats.ru/_avtoreferats/01007498640.pdf, last accessed 28.11.2020.
6. Nizomutdinov, B., Tropnikov, A., Uglova A.: Application of social networks user' digital fingerprints to predict their information image. Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance (ICEGOV2020), Athens, Greece, April 1–3, 2020. ACM New York, NY, USA (2020), https://doi.org/10.1145/3428502.3428635, last accessed 28.11.2020.
7. Nizomutdinov, B.A., Tropnikov, A.S., Uglova, A.B. Razrabotka prognosticheskoi modeli informatsionnogo obraza polzovatelia s primeneniem avtomatizirovannykh sredstv obrabotki dannykh iz sotsialnykh setei. Nauchnyi servis v seti Internet, (21). 532–540 (2019), https://www.elibrary.ru/item.asp?id=41256649, last accessed 28.11.2020.
8. Pianesi, F., Mana, N., Cappelletti, A., Lepri, B., Zancanaro, M.: Multimodal recognition of personality traits in social interactions. Proc of the 10th International Conference on Multimodal Interfaces, 53–60 (2008), https://doi.org/10.1145/1452392.1452404, last accessed 28.11.2020.
9. Settanni, M., Azucar, D., Marengo, D.: Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. Personality and Individual Differences, 150–159 (2017), https://doi.org/10.1016/j.paid.2017.12.018, last accessed 28.11.2020.
10. Soloveva, L.N. Tsifrovaia identichnost kak novyi vid identichnosti cheloveka informatsionnoi epokhi. Obshchestvo: filosofiia, istoriia, kultura (2018), doi: 10.24158/fik.2018.12.6, last accessed 28.11.2020.
11. Tropnikov, A., Uglova, A., Nizomutdinov, B.: Development of a prognostic model of the user's information image using automated tools for processing data from social networks. Communications in Computer and Information Science, 1038, 405–413 (2019), https://link.springer.com/chapter/10.1007%2F978-3-030-37858-5_33, last accessed 28.11.2020
12. Uglova, A.B., Koroleva, N.N., Bogdanovskaia, I.M., Lugovaia, V.F.: Strategii virtualnoi samoprezentatsii sovremennykh rossiiskikh uchitelei v sotsialnykh setiakh. Pisma v Emissiia.Offlain: elektronnyi nauchnyi zhurnal, (10), 2773 (2019), http://www.emissia.org/offline/2019/2773.htm, last accessed 28.11.2020.
13. Waterloo, S.F., Baumgartner, S.E., Peter, J.: Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. SAGE, (20), 1–19 (2017), https://doi.org/10.1177/1461444817707349, last accessed 28.11.2020.
14. Voiskunskii, A., Evdokimenko, A., Fedunina, N.: Setevaia i realnaia identichnost: sravnitelnoe issledovanie. Psikhologiia. Zhurnal Vysshei shkoly ekonomiki, 10(2), 98–121 (2013), https://publications.hse.ru/articles/101407407, last accessed 28.11.2020.

15. Vartanova, E.L.: The media and the individual: economic and psychological interrelations. Psychology in Russia: State of the Art, 6(1), 110–118 (2013), http://psychologyinrussia.com/volumes/index.php?article=2089, last accessed 28.11.2020.
16. Vapnik, V.N.: An Overview of Statistical Learning Theory. Neural Networks, IEEE Transactions on., 10(5), 988–999 (1999), https://doi.org/10.1109/72.788640, last accessed 28.11.2020.
17. Voiskounsky, A.Ye.: Psychology of computerize ation as a step towards the development of cyberpsychology. Psychology in Russia: State of the Art., 6(4), 150–159 (2013), http://psychologyinrussia.com/volumes/pdf/2013_4/2013_4_150-159.Pdf, last accessed 28.11.2020.
18. Xenos, S., Ryan, T.: Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. Computers in Human Behavior, 27(5), 1658–1664 (2011), https://doi.org/10.1016/j.chb.2011.02.004, last accessed 28.11.2020.