Thematic Modeling of Professional Communication Practices of Helping Specialists

A. B. Uglova[1], I. M. Bogdanovskaya[1], and B. A. Nizomutdinov[2(✉)]

[1] Institute of Psychology, The Herzen State Pedagogical University of Russia, Saint Petersburg, Russia
anna.uglova@list.ru, ibogdanovs@herzen.spb.ru

[2] Institute of Design and Urban Studies, ITMO University, Saint Petersburg, Russia
boris@itmo.ru

Abstract. The paper examines the content and structure of professional communicative practices of helping specialists in digital professional communities, and also identifies psychological influence techniques used to manage interaction with the audience. A platform with profiles of psychologists offering their services was selected for the study. With the help of parsing methods, all the specialists' questionnaires were saved. Further, thematic modeling was carried out using LDA methods. 10 main topics and 55 unique terms describing the content of the professional discourse of helping specialists were identified. As a result, the components of the communicative strategy of using digital resources were evaluated: the presence of professional vocabulary, basic terminology, descriptions of transformational techniques, the presence of a product image, an image of the result, a description of the target audience, the presence of suggestive manipulative technologies. The model of monitoring the professional digital community as a whole, the possibilities of the platform for implementing elements of the communicative strategy of specialists were also described..

Keywords: parsing · data collection · mathematical modeling · psychology · specialists

1 Introduction

At a fundamental level, in the era of the fourth industrial revolution, there are issues of changing the healthcare system, the possibility of introducing cyber-physical systems into the care system, and the impact of digitalization on the quality of psychologists' work. It becomes important to study information technologies that are used to connect helping specialists with clients, as well as to transfer knowledge and psychological online help [1].

Within the framework of the UNESCO initiative related to the modernization of the health saving system, an important stage in rethinking the system of psychological assistance is the problem of analysis, evaluation and design of constructive communication practices in a virtual environment.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023, LNCS 13957, pp. 441–452, 2023.
https://doi.org/10.1007/978-3-031-36808-0_31

The relevance of the study of communication practices in a virtual environment is associated with the need to organize a successful dialogue between specialists and clients. The search for a helping specialist requires a high informational and psychological culture from the user. The client's choice is determined by the information that the assisting specialist presents on the web and popular information trends that are created by the digital professional community as a whole. The success of the interaction will depend on how well the user will understand the communication strategies and tactics of specialists [2].

In our study, we define professional communication practices as a complex multi-dimensional process of establishing and developing contacts generated by the needs of people's joint life activity in a virtual environment [3].

Communication practices in a virtual environment include: 1) communication actors, as well as technical objects and socio-cultural artifacts (computer intellectual agents, content, etc.), 2) are associated with the search and exchange of information, ensuring the social community of communicants while preserving their individuality, 3) function in accordance with the rules of a particular network community and 4) are mediated by a number of principles of digital media [4].

To improve information and psychological literacy, it is important to study the leading communicative practices of helping specialists in the digital environment [5].

Communication practices in digital media have a number of features that distinguish them from traditional social practices. These include virtuality, textual character, interactivity, hypertextuality, uncertainty of spatial and temporal localization, creativity [6]. At the same time, communication practices in the virtual space include such characteristics of digital media as: programmability (the ability to make changes to the mode of communication using technical means), modularity (the ability to work with individual elements of a web page, the use of links, etc.), automation (standardization of many functions and elements of communication), variability (the ability to create multiple copies of a single object, which will be modified somewhat if desired by the user), hypermedia (the presence of non-text elements in hypertext), scalability (the ability to choose a simplified or full version of the site), periodic updating, transcoding (mutual projection of the communication process in the cultural and computer, technological layers) [7].

Since these characteristics add a sense of engagement and personal communication to communication, they are able to increase customer loyalty and reduce the criticality of his attitude to the information consumed [8]. For this reason, specific cyber risks appear, which are important to study for the organization of constructive Internet interaction [9].

An additional aspect of the relevance of the study of communicative practices in the virtual space is their, in many cases, openness and accessibility for observation. For this reason, cyberspace is currently one of the main platforms for the study of man. Mastering communicative practices in a virtual environment, the user learns and generates new values, norms and patterns of behavior, forms a system of actions that allow adequately perceiving and transmitting information to achieve both individually significant and group goals [10]. These features allow us to consider communicative practices in the virtual space, including as a methodological basis for studying social phenomena and processes.

In connection with the above, we would like to present 2 research questions.

RQ1. What is the specificity of the content and structure of professional communication practices of helping specialists in digital professional communities?

RQ2. What methods of psychological influence are used to manage interaction with the audience?

Currently, for online search of psychologists and psychotherapists in Russia, either specialized specialist recruitment services or websites for posting profiles of freelance specialists are mainly used. The choice to study the B17 platform was due to the fact that at the moment it is one of the oldest and most popular services for posting questionnaires of specialists of helping professions.

2 Data Collection

As previously noted, in our work we studied information about psychologists and helping specialists through a resume on the B17 Website. This is a social network that allows users to create personal profiles, write blog posts, share photos and videos, and communicate with other users on the topic of psychology. The website features a variety of topics including relationships, lifestyle, entertainment, and self-expression. The name of the Site "B17" comes from the phrase "be yourself at 17", which reflects the site's emphasis on self-expression and individuality. The website also has a forum where users can discuss various topics and ask for advice from other users, as well as a messaging system for private communication between users. We studied only the open part - the profiles of specialists, that is, only the information that users themselves openly publish about themselves.

The parsing method was used to collect the data. Data parsing is the process of extracting information from structured or unstructured data, which may be different in format and source. With the help of parsing, you can automatically extract and process information from various sources, such as web pages, databases, document files, etc.

Various methods and technologies are used for parsing, including regular expressions, machine learning algorithms, as well as special programming tools and libraries. The result of parsing can be data in a format that is convenient for further use for various purposes.

The surest way is to collect data through the API. An API (Application Programming Interface) is a set of instructions and data structures that programmers can use to create applications. The API provides a standardized way to interact between different applications, allowing them to exchange information and functionality. Interaction with the API can be carried out via communication protocols, such as HTTP or TCP/IP. As a rule, APIs are used to access functions and services of remote systems or external resources, for example, databases, calendars, social networks, online stores, etc. We examined the B17 site for the presence of an API, and did not find such an interface. Therefore, it was decided to use data parsing methods using a crawler.

A crawler for parsing (also called a web scraper or web spider) is a program that automatically crawls websites, extracts the necessary information and saves it in a convenient format, for example, in a database or file. The crawler starts working from a certain start page, and then follows links to other pages (depending on the set rules and logic), extracts the page content, processes it and saves it in the appropriate format. The

crawler is used to automate the extraction of information from websites, which speeds up and simplifies the processing of a large amount of data.

However, it should be remembered that the use of crawlers may violate the rules of use of sites (depending on the specific case), therefore, in order to use a crawler, it is necessary to comply with the rules of a particular website and legislation. We have studied the terms of use of this website and found no contradictions. It is also worth noting that we used all the collected data only for scientific research, we do not use this data for commercial purposes.

All the collected data were additionally depersonalized. For this task, we conducted a depersonalization stage of the data. Depersonalization is the process of deleting or replacing personal data that can identify an individual in order to protect privacy and confidentiality. The depersonalization procedure allows you to hide or replace sensitive information, while preserving the value of the data for analysis and other purposes.

All collected information was preprocessed to remove any proper names. To solve this problem, a script was written using the Natasha library. Natasha is a free library for natural language processing (NLP) for texts developed in Python. It has an open source code and is used for analyzing texts in natural Russian. As a result, all the collected data were depersonalized and did not contain contacts, names and surnames. All this was necessary to comply with the legislation on the processing of personal data.

The Phython language was chosen for writing the parser - it is widely used for writing parsers due to its rich libraries and ease of use. Additionally, 2 libraries were used - Selenium and Scrapy. Python Selenium is a Python library that allows you to automate web browsers using WebDriver. Selenium also allows you to record and replay actions in the browser, which can greatly simplify the process of testing web applications or automating routine actions.

Python Scrapy is a Python framework for collecting data from websites. It provides a powerful set of tools for automatic data extraction, tracking changes on pages and automating work with websites. Scrapy makes it easy to create and manage web scrapers using Python. It provides a set of convenient and flexible tools for data extraction and processing, including multithreading, proxy support, monitoring changes on pages and other functions.

Scrapy provides automatic page navigation, interactive form processing, AJAX parsing and other functions for working with complex websites. In addition, it has a large number of additional functions, such as searching for elements on the page, using cookies and managing browser windows. This makes it one of the most popular tools for automating web browsers. The general scheme of work is shown in Fig. 1.

The result of data collection was the MySQL database. The following parameters are stored in the database:

- Specialty
- City
- Number of publications
- Number of reviews
- Number of examples of consultations
- Number of trainings
- Number of online consultations

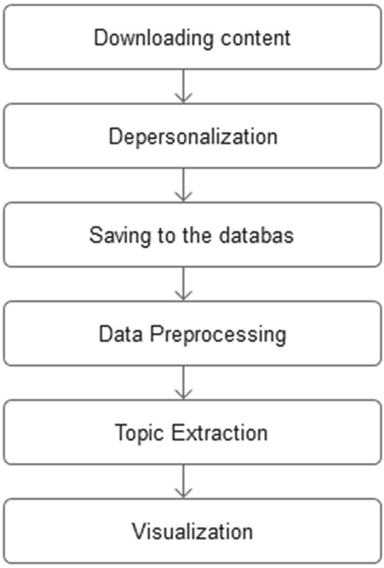

Fig. 1. General scheme of data collection and processing

- Full text description of the specialist

The final database contains 2224 records. As a result, we managed to collect a database of psychologists from the largest cities for further processing.

3 Data Processing

The next step was data processing. In our work, we have carried out intelligent data processing for the descriptions of psychologists that are published in profiles. To identify the main areas in which specialists work. To solve this problem, we conducted thematic modeling based on full-text descriptions of specialists. 10 main topics were identified.

Orange Text Mining software was selected for data processing. It is a free Python tool used to analyze text data. It provides a set of tools that facilitate the analysis of a large volume of text data, including text classification, clustering, visualization, keyword extraction, and others. It has a rich set of tools that allow you to analyze texts, starting from the basic analysis of the frequency of words and ending with the combination of various machine learning methods for the classification and clustering of texts. Another interesting feature of Orange Text Mining is the ability to visualize text data. It provides tools that allow you to create visualizations, such as a tag cloud, heat maps, a graph using the TSNE method, and others [11]. One of the most useful features of Orange Text Mining is the ability to extract keywords. It contains a set of tools that allow you to process text documents and extract keywords that can help in analyzing text data. In general, Orange Text Mining is a powerful and flexible tool for text data analysis, which can be used for both small and large projects in various fields, such as marketing, science and technology related to text data analysis.

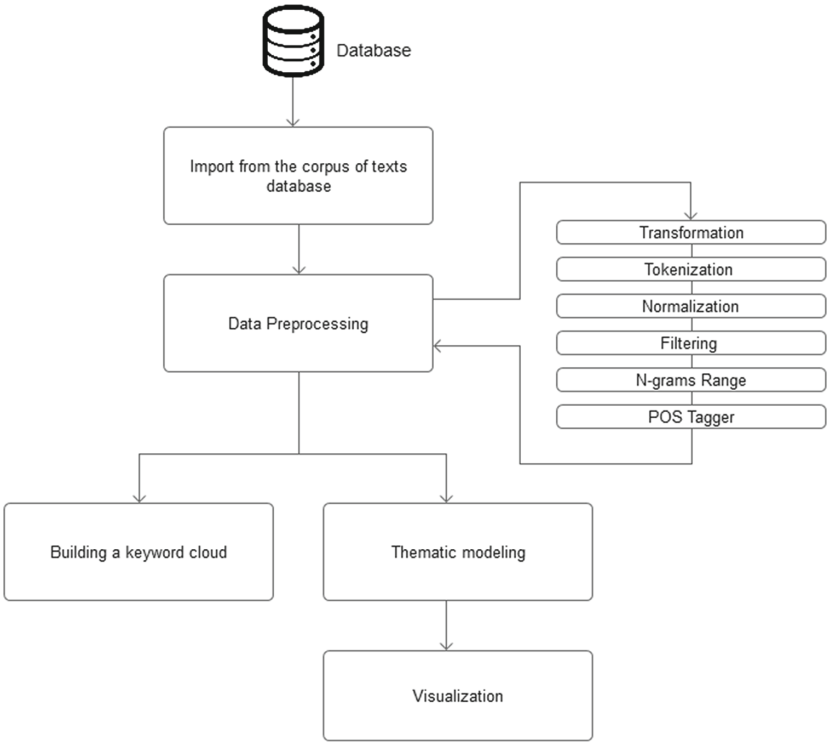

Fig. 2. General text processing scheme for thematic modeling

The general scheme of thematic modeling is shown in Fig. 2.
In our work, we used the following tools for text processing:

- Preliminary preparation of the text. Text preprocessing breaks your text into smaller units (tokens), filters them, starts normalization (stemming, lemmatization), creates n-grams and marks tokens with part-of-speech labels. Additionally, a dictionary with negative words was prepared to remove unnecessary information, such as emojis and other information.
- Building a keyword cloud. Word Cloud displays tokens in the corpus, their size indicates the frequency of a word in the corpus or the average number of words in a packet when the functions of the word packet are at the widget input. The words are listed in the widget by their frequency (weight). The widget outputs documents containing the selected tokens from word cloud.
- Thematic modeling. Thematic modeling identifies abstract topics in the corpus based on groups of words found in each document and their corresponding frequency. A document usually contains several topics in different proportions, so the widget also reports the weight of topics for each document. Thematic modeling can be carried

out using various methods. Using a hidden Dirichlet distribution, hidden semantic indexing, or a hierarchical Dirichlet process.

Among the methods of thematic modeling, many researchers still use LDA, a generative model that observes the frequency of words in the corpus and iteratively builds a thematic model for a given number of topics [12]. We settled on the LDA method, because in our context it showed the optimal result.

For thematic modeling in Orange Text Mining, the Gensim library is used. Gensim is an open source library for unsupervised thematic modeling, document indexing, similarity search and other natural language processing functions using modern statistical machine learning. The result of the processing were 10 topics described by keywords, as well as the possibility of visualization. Each cluster was interpreted by our team, the results of cluster processing are given in the next chapter.

4 Results

Professional communication practices are being transformed in the digital environment, acquiring new functions and forms. However, the structure of the communicative strategy remains classical, which allows us to evaluate the main components of the content: why the material is published, what the professional wants to say about himself, to whom the self-presentation is addressed, how the final result of the work is seen, what vocabulary and forms of speech are used, how grammatical forms are used that can affect the process of perception and understanding of information by the reader (potential client).

At the first stage, we conducted a semantic analysis of the uploaded information and identified 55 unique terms describing the content of the professional discourse of helping specialists (Table 1).

Assessing the content of the professional digital discourse of helping specialists, we can say that they use a wide range of vocabulary that has both direct and intentional meaning.

The general psychological vocabulary is widely used to indicate that a specialist belongs to a certain professional group. It is worth noting the presence of both medical and psychological terminology, which indicates the presence of specialists of different profiles in the community.

The basic directions of psychotherapy and counseling, a wide range of approaches and techniques that are used for treatment, improvement of mental health and personal development in individual work with a client are described. As a result of the analysis, the basic areas of work were identified: psychoanalytic approach, gestalt therapy, cognitive behavioral therapy. As well as the most popular methods of work: positive psychotherapy, Hellinger placement, art therapy, AST (acceptance and commitment therapy).

The communication strategy also provides a description of the main object of the work - the image of a psychological request. The request to a psychologist is primarily related to the problems faced by clients (the sphere of communication, self-realization, health, etc.). The creation and presentation of a certain “standard” request affects the process of reflection and sets a trend in self-understanding among customers. Specialists primarily appeal to the spheres of relationships, work, and education. Separately, it is

Table 1. The content of the professional discourse of helping specialists

General psychological vocabulary			
Therapy	Psychotherapy	Consultation	Example
Psychology	Counseling	Help	Course
Transfer	Psychotherapeutic	Medical	Scheme
Basic directions of psychotherapy and counseling			
Gestalt	CBT (Cognitive Behavioral Approach	Psychoanalytic	ACT (Acceptance and commitment therapy)
Psychoanalysis	Positive Psychotherapy	Arrangements by Hellinger	Art Therapy
The image of a psychological request			
Situation	Condition	Case	Problem
Life	Relationship	Work	Education
Speech	Speech	Difficulty	Weight loss
Image of the result			
Success	Happy		
Psychological transformation techniques			
Mental changes	Emotional changes	Positive changes	
Responsibility	Development	Acceptance	
Target audience			
Children	Child	Client	
Intention of commitment (commission)			
Be Able	To Want	To Do	
Find	To Know	Work out	

worth noting the popularity of the topics of weight loss and speech development as basic psychological trends.

A place is given to the description of the idealized image of the result (happiness and success). The image of the result as a mental picture of what we want to achieve is a visualization of the desired result. In this case, the desired result is described quite abstractly, which on the one hand makes it possible to attract more customers, and on the other hand creates conflicting expectations in relation to specialists.

The description of specialists presents psychological transformational techniques for achieving results, which are designed to help change behavior, thinking or emotional state, achieve personal growth, as well as overcome various problems and conflicts. The greatest emphasis is on the techniques of taking responsibility therapy, which is associated with the leading areas of therapy. Suggestive intentional statements (commissives) are also actively used: I can, want, do, find, know, try, encouraging active actions of clients.

The target audience that experts plan to interest is described rather sparingly. There is a separate emphasis on working with children. Adults are not described in a differentiated way, messages are shared, without personal addressing, which reduces the effectiveness of communication.

At the second stage, using thematic modeling, we evaluated the structure of the communicative strategy of helping specialists (Table 2).

Table 2. Structure of the communication strategy of helping specialists

Clusters	Tags
Attributes of medical psychology, psychotherapy	Psychotherapy, psychology, psychotherapy, positive, psychotherapeutic, medical, positive, education, clinical, center
Cognitive behavioral therapy of taking responsibility	Cognitively, acceptance, client, transference, AST, case, responsibility, therapy
Cognitive-behavioral approach to working with weight loss	CBT, course, condition, success, cognitive, psychotherapy, therapy, scheme, weight loss, gestalt
Gestalt therapy of child-parent relations	Example, gestalt, counseling, children, psychology, emotionally, work, psychotherapy, therapy, relationships
Techniques and tasks of gestalt therapy	Happiness, games, find, you know, work, do, gestalt, life
Techniques and tasks of art therapy	Life, relationships, therapy, art, problems, situations, stage, help, development, want
Psychoanalytic work with children	Psychoanalysis, psychotherapy, psychoanalytic, psychoanalytic, work, children, counseling, psychotherapy, psychoanalysis, psychology
Integrative approach in working with the sphere of relations	Gestalt, help, relationships, CBT, min, therapy, consultation
Positive psychotherapy and Hellinger placement	Dynamic, example, placement, positive, consultation, psychotherapy, psychotherapeutic, mental
The technique of the Hellinger method of arrangements	Speech, speech, speech, speech, a, can, difficulties, various, constellations

Based on the results of the structural analysis, it is possible to identify the main topics and describe the problems with which specialists are ready to work in various directions. For example, it is proposed to work with the topic of weight loss within the framework of a cognitive-behavioral approach, and with children, specialists offer therapy within the framework of psychodynamic and gestalt approaches. It can also be noted that the psychodynamic direction, including art therapy, psychoanalysis, the

method of placement, positive psychotherapy, is the most popular and is most represented in professional discourse. With the help of thematic modeling, it is also possible to evaluate the structuring of the description of individual directions, the inclusion of the user's image, the targeting of self-presentation, the description of techniques and the result.

As a result of the analysis, the following constructive and destructive characteristics of professional communication on digital platforms can be described:

- targeted openness, which is characterized by the absence of a clearly defined audience and focus on the description of the problem field;
- genre diffuseness associated with the combination of often contradictory directions of assistance in one message;
- idealization of the image of the result of the work, forming unrealistic expectations of the audience;
- the use of suggestive statements (commis-sives) to encourage the audience to communicate.

The leading characteristics of professional psychological discourse indicate the integration of its types as professional-practical and advertising-commercial, which correlates with the main objectives of such platforms, but may contradict the ethical principles of helping specialists, which needs further study.

At the same time, the communication strategies of specialists in the professional community are mediated by such characteristics of digital media as the ability to work with individual elements of a web page, the use of links, standardization of many functions and elements of communication, hypermedia (the presence of non-text elements in hypertext), periodic updating, etc. Thus, with the help of thematic modeling, it is possible to evaluate not only the work of individual specialists, but also to monitor the professional digital community as a whole. Based on the results of the analysis, it can be said that the digital platform under study is focused on the presentation of the following elements of the communication strategy:

- Analysis of the goals and objectives of psychological assistance, through an active description of the issues;
- Self-presentation of professional achievements and skills of specialists.

However, almost no attention is paid to the description of the target audience and segmentation of consumers of services to improve targeted communication. Also, a clearer description of the result of work can become a point of development, in order to increase the understanding of customers of what they can get as a result of work and increase the desire to cover current needs with the help of the services offered.

Discussion.

The results obtained allow us to conclude that we have received answers to the research questions posed. There is a specificity of the content and structure of

professional communicative practices of helping specialists in digital professional communities:

- the community of helping specialists has absorbed the basic properties of virtual communities associated with hypertextuality (the use of multimodal messages, the intersection of social networks, the use of different information channels, etc.), targeted openness, crowdsourcing, the absence of geographical boundaries, etc.
- the virtual community has retained offline functions: the possibility of self-presentation, exchange of professional information, etc.
- the virtual community has created new properties of the professional environment of psychologists: genre diffuseness, changing the ethical norms of self-presentation of services, expanding the space for communication with patients, creating new conditions for seeking psychological help.

These conclusions correlate with the results of research by Wu H. et al., which indicate that in online interaction, clients show more trust in psychologists than in offline interaction, based on the assessment of the resource and the community as a whole [13]. We can say that the question of the ethics and usefulness of such communities for customers remains open and needs further research.

5 Conclusions

In conclusion, I would like to say that with the help of thematic modeling, it is possible to assess the content and structure of the areas of work of the helping specialists. It also became possible to evaluate the communicative strategy of using digital resources: the presence of professional vocabulary, basic terminology, descriptions of transformational techniques, the presence of a product image, an image of the result, a description of the target audience, the presence of suggestive manipulative technologies. This model can be used to evaluate various professional communities.

Also, with the help of thematic modeling, it is possible to evaluate not only the work of individual specialists, but also to monitor the professional digital community as a whole, the capabilities of the platform for the implementation of all elements of the communicative strategy.

In the continuation of this study, we plan to study the regional specifics of professional communication of helping specialists to refine the monitoring model taking into account socio-demographic and territorial specifics.

The advantage of this method is that it can be scaled for any platform containing specialist questionnaires. In our study, we studied one site, however, the parser can be configured to collect information from other sites, and collect data. Further thematic analysis will not require refinement and the same algorithm will be used.

6 Conclusions

The research was carried out at the expense of the grant of the Russian Science Foundation No. 22–78-10047, https://rscf.ru/project/22-78-10047 / “Constructive and destructive communicative practices of specialists of helping professions in digital media”.

References

1. Wu, H., Dan, J., Evans, R.: Building patient confidence in psychologists in online mental health communities. Data Sci. Manage. **5**(1), 21–27 (2022)
2. Heather, H.J.: Dialogic communication in the context of healthcare: a case study of Kaiser Permanente practice in social networks. Overv. Publ. Relat. **40**(5), 856–858 (2014). https://doi.org/10.1016/j.pubrev.2014.07.007
3. Warren, M.: "Just spoke to …": The types and directionality of intertextuality in professional discourse. English Specific Purp. **32**(1), 12–24 (2013). https://doi.org/10.1016/j.esp.2012.07.001
4. Radkevich A.L.: Social Internet practices as an object of sociological analysis. Knowledge. Understanding. Skill: information humanitarian portal. Sociology. № 3 (2009). http://www.zpu-journal.ru/e-zpu/2009/3/Radkevich
5. Karasne, R., Al-Azzam, S., Muflikh, S., Suda, O., Khawamde, S., Khader, Y.: The influence of the media on the formation of knowledge, awareness of risks and communication practices in connection with the COVID-19 pandemic among pharmacists. Res. Field Soc. Admin. Pharm. **17**(1), 1897–1902 (2021). https://doi.org/10.1016/j.sapharm.2020.04.027
6. Marres, N.: Digital Sociology. Rethinking the social. Research, p. 232p. Polity Books, Cambridge (2017)
7. Manovich L. The Language of New Media, 400 p. Moscow (2018)
8. Rose, S., Fandel, D., Dibley, A.: Information exchange is the name of the game: a study of the role of communication practices in social networks in relations with B2B clients in the natural sciences industry. Ind. Mark. Manage. **93**, 52–62 (2021)
9. Bogdanovskaya, I.M., Koroleva, N.N., Uglova, A.B., Petrova, Y.: The role of personal characteristics in the formation of problematic Internet use among high school students. Prosp. Sci. Educ. **54**(6), 271–284 (2021). https://doi.org/10.32744/pse.2021.6.18
10. Zotov V.V., Lysenko V.A.: Communicative practices as a theoretical construct of the study of society. Theory and practice of social development. No.3 (2010). https://cyberleninka.ru/article/n/kommunikativnye-praktiki-kak-teoreticheskiy-konstrukt-izucheniya-obschestva. Accessed 20 Mar 2023
11. Lamba, M., Madhusudhan, M.: Tools and techniques for text mining and visualization. In: Lamba, M., Madhusudhan, M. (eds.) Text Mining for Information Professionals: An Uncharted Territory, pp. 295–318. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-030-85085-2_10
12. Sievert, C., Shirley, K.E.: LDAvis: a method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, pp. 63–70, Baltimore, Maryland, USA (2014)
13. Wu, H., Deng, Z., Evans, R.: Building patients' trust in psychologists in online mental health communities. Data Sci. Manage. **5**(1), 21–27 (2022)
14. Jiang, Y., Hu, C., Xiao, T., Zhang, C., Zhu, J.: Improved differentiable architecture search for language modeling and named entity recognition. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 3585–3590 (2019). https://doi.org/10.18653/v1/D19-1367
15. Dieng, A.B., Ruiz, F.J.R., Blei, D.M.: Topic modeling in embedding space. Trans. Assoc. Comput. Linguist. **8**, 439–453 (2020). https://doi.org/10.1162/tacl_a_00325
16. Wolf, T., Debut, L., Sanh, V.: Transformers: state-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 38–45 (2020). https://aclanthology.org/2020.emnlp-demos.6
17. Hwang, C.L., Lai, Y.J., Liu, T.Y.: A new approach for multiple objective decision making. Comput. Oper. Res. **20**(8), 889–899 (1993). https://doi.org/10.1016/0305-0548(93)90109-v