Technology and digitalization: organization of scientific research in the magistracy

O V Kononova1 and D E Prokudin1,2

1 ITMO University, St Petersburg, Russia
2 St Petersburg State University, St Petersburg, Russia

E-mail: ovkononova@itmo.ru, hogben.young@gmail.com

Abstract. The problem of effective search, explication, and analysis of contextual knowledge is associated with an ever-increasing volume of available information, its non-formalization and poor structuredness, a high rate of updating of scientific knowledge, the polysemy of the terminological base used, the transfer of terms from one subject area to another without the necessary interpretation and adaptation. The existing tools for searching, explicating and analysing contextual knowledge are of little demand due to the lack of meaningful information both about the tools themselves and about the methods and algorithms for their application in science and practice, as well as the limited set of services provided by analytical software. The educational-methodical complex “Technologies of data extraction and mining in scientific research” is aimed at the formation of research and analytical competencies of undergraduates. The results of mastering the educational product will allow undergraduates to consciously apply in research practice an integrated approach (synthetic method) to the search and analysis of contextual knowledge, to use analytical software and environments with built-in services for explication, clustering, and statistical processing of scientific texts on the basis of this approach.

1. Introduction
In the Russian scientific, socio-political and socio-economic discourse, the concepts of “digital technologies”, “digital platforms” and “digital society”, “digital space” appeared relatively recently, replacing the familiar, well-established terms “information technologies”, “information platforms” and “information resources”, “information society.” Moreover, the term “digitalization” of the economy, production, the public sector, and other spheres and industries is being used more often or on a par with computerization in business circles. What is behind the terminology change? A change in the technological order, lifestyle or a tribute to fashion and a consequence of wide media and political rhetoric, which will “sink into oblivion” along with the completion of national, state, regional and related programs and projects oversaturated with this terminology. This is a reason for separate reflection and research.

So, according to a number of researchers, the modern society, called digital, is characterized by a change in the technological order, which corresponds to another term – “Industry 4.0”, introduced into scientific circulation in 2011. The concept of “Industry 4.0” is interpreted as a transition to fully automated digital production controlled by intelligent systems in real time in constant interaction with the external environment, going beyond the boundaries of one enterprise, with the prospect of uniting into a global industrial network of things and services [4]. The fourth industrial revolution changes the
way of doing business not only in industry, but also in all sectors of the economy and social and political life [5, 6, 7].

Current scientific analytics [3, 8, 12, 13, 14] have shown that both throughout the world and in Russia, the interest of the scientific community in the trends in the development of the digital economy is growing. Moreover, the dynamics of the Russian scientific discourse is in agreement with global trends. And the explosive growth in the interest of the Russian scientific community in the problems of developing the digital economy since 2017 is explained by the beginning of the digital economy development at the state level. This is clearly shown by the dynamics of the information flow in the digital economy in the socio-political discourse according to data from the Russian mass media (figure 1). Such a rapid development of the concept of the digital economy, both throughout the world and in Russia, leads to the transformation of various types of human activities, requiring specialists to possess new competencies. This, in turn, leads to the development of training programs for specialists, taking into account new challenges and needs. These programs require both a change in approaches to the organization of the educational process, and a rethinking of educational tasks and content components of programs. From our point of view, first of all, master's programs should and are undergoing change as shorter, which means mobile – designed to flexibly and quickly respond to the needs of a society moving along the path of "digitalization", and saturated with research components, ending with the defense of a master's thesis. The importance of the research component of master's programs, especially in the field of ICT, is confirmed by the emergence on the labour market of the already quite widespread and demanded professions, for example, data analyst and data scientist.

Figure 1. Dynamics of changes in the flow of scientific publications and the number of publications in the media. Query: “digital economy”

As part of the Potanin Foundation's support program for master's programs, ITMO University is implementing a project to develop an educational and methodological complex (EMC) “Technologies for Data Extraction and Data Mining in Scientific Research” aimed at developing the knowledge and skills of applying modern information and communication technologies (ICT) in their research and
design activities. EMC assumes the complex use of technologies for the search, extraction and analysis of scientific knowledge. This approach makes it possible to more efficiently search, extract and process information from open databases and scientific sources, better navigate the developing interdisciplinary fields, in which the terminological base has not yet been established.

The project is aimed, first of all, at supporting master's programs similar to the program “Digital technologies of a smart city”. This program was developed and implemented at the Institute of Design and Urbanism (IDU) at ITMO University. It meets the needs of the public sector in ICT, the main trends and prospects for the development of ICT in the public sector, approaches to its digitalization, the vector of the development of informatization of the activities of public authorities and urban management, taking into account digital initiatives and smart city projects in Russia.

The use of techniques based on intelligent data processing technologies, systems of advanced full-text and multimodal search, methods and tools for extracting contextual knowledge with the simultaneous mastery of a wide range of analytical tools will allow undergraduates, graduate students and other categories of researchers to increase the effectiveness of their analytical and research activities at the university. The relevance of the educational product being created is determined by its compliance with the needs of a modern digital society for the consumer qualities of information, methods and tools for obtaining and analysing it, and, accordingly, for the process and results of training specialists. The demand for the product by the university and educational community is due to the need to timely take into account the trends in the development of the socio-economic sphere in the development and modernization of curricula and training programs for undergraduates, where future masters will have to apply the acquired professional knowledge and competencies.

The educational-methodical complex includes:

- study guide and terminological dictionary;
- recommendations for working with the Scientific Electronic Library (http://elibrary.ru) information system, other scientific information resources and the media in integration with independent analytical systems;
- recommendations for the use of classes of analytical systems and analytical environments T-Libra, Voyant-tools, Tropes High Performance Text Analysis, Sketch Engine, BigARTM, Mallet, machine learning systems, etc.

Part of the teaching materials are open access resources:

- open access repository for posting the results of practical tasks in the form of thematic collections of contexts and thesauri (based on the open source software DSpace);
- aggregator of metadata of scientific publications and other information materials (based on open source software Open Harvester Systems);
- an electronic annotated catalogue of computer systems to support scientific research and analysis of contextual knowledge, highlight and explicate scientific content.

The methodological and substantive novelty of the proposed teaching materials lies in the use of an integrated approach (synthetic method), independent of the toolkit, in R&D of undergraduates to solving problems: highlighting and explicating scientific content, searching and explicating contextual knowledge, data mining; applying typology, semantic mapping and structured description of contextual knowledge. The innovativeness of the approach has been proven by testing on a wide range of analytical software, supported by the scientific community: publications and reports at international conferences [3, 9, 10, 11].
2. EMC “Technologies for data extraction and mining in scientific research”

2.1. On research work as part of the educational process of the university

Organization of research work at a university is a complex activity, the success of which today is determined not so much by educational standards as by the digital technologies used to implement them. The standards set for higher education such tasks as analysis and generalization of the results of interdisciplinary research, processing of large arrays of heterogeneous data with the subsequent preparation of scientific publications on research topics. When searching for information, a huge amount of source material is accumulated, which must be preserved and in the future be able to process and analyse it. Therefore, the skills of searching, analysing and selecting useful information are among the most relevant for researchers, teachers and students, requiring them to have certain competencies and skills [1, 2].

Master’s program “Applied Informatics. Digital Technologies of the Smart City” ITMO provides training in the field of information technologies for the city and state sectors of the economy. Professionally oriented training of graduates is ensured through the inclusion of undergraduates in real research and development on the development of the city at the intersection of interaction between science, business and government. The program is focused on research and project activities, which determines the topic of research and a high degree of student involvement in urban projects and initiatives.

The proposed educational and methodological complex (EMC) complies with the ITMO educational standard “Applied Informatics” and the content of the discipline “Technologies for data extraction and mining in scientific research”. The EMC offers an organizational and meaningful rethinking of the research work of undergraduates based on the use of contextual search and data mining technologies, which makes it possible to identify and correlate contexts of textual and non-textual modality with areas of knowledge and the object of research, interpret trends that demonstrate the dynamics of the development of interdisciplinary research areas.

The teaching materials also pay attention to the formation of skills for the rational selection of information sources, conducting rational thematic search in various information retrieval systems, using modern tools and technologies for presenting the results of research activities. Integration of teaching materials into the educational process will provide graduate students with methods and tools to achieve practice-oriented results that are in demand by government agencies and the business community.

2.2. Synthetic method

Reconsidering the role of interdisciplinary scientific fields and the applicability of traditional scientific methods in interdisciplinary research, issues related to the explication and subsequent analysis of contextual knowledge in order to identify new promising research areas are widely covered in scientific publications. Contextual knowledge is usually understood as the ability to correctly “read” the context, extract and interpret professionally relevant information through contextual search. As a result, both traditional scientific approaches and modern methods and tools of Digital Humanities, data mining were taken into account in the structure and content of the teaching materials.

The basis for the preparation of undergraduates using the developed EMC is the integrated approach (synthetic method) developed by the authors to search and highlight contexts in interdisciplinary research and build on its basis a methodology for analysing and explicating contexts, a method for constructing and interpreting trends. The approach is based on the use of information and communication technologies to research the development of various subject areas and practices of human activity through the study of the formation and development of their conceptual and terminological apparatus. The synthetic method is aimed at extracting contextual knowledge from unstructured or semi-structured information resources and allows, through explication and mapping, to form collections of relevant fragments (thematic contexts). The application of the synthetic method does not depend on the choice of specific analytical information systems, which provides flexibility in its use and availability of application depending on the capabilities of research teams.
The approach used in the study is consistent with the methods adopted in this kind of research. However, its peculiarity lies both in taking into account and integrating various methods used, and in its own methods, which include:

- refusal to study a thematic sample of highly cited scientific journals with a high impact factor in favour of considering a wider range of publications from thematically different editions;
- synthesis of various methods, integrated coverage of research tools, and varying the sequence of application of technologies for search, selection, explication, and analysis of contextual knowledge, depending on the initial conditions and characteristics of particular research;
- implementation of the possibility of creating and further replenishing individual thematic collections of thematic contexts based on clustering of query results.

The use of the synthetic method in the framework of the EMC is aimed at developing skills for the integrated use of modern information and communication technologies in solving problems of searching, extracting, explicating, and analysing both scientific and professionally significant information. In general, through the development of the synthetic method and the fulfilment of relevant educational tasks, the professional competencies of the researcher and analyst are formed, including for solving applied problems using scientific methods and tools.

2.3. IS catalogue
The developed and developed integrated approach implies the integrated use of electronic library services and analytical information systems, such as T-Libra, Voyant-tools, Tropes High Performance Text Analysis, Sketch Engine, BigARTM, Mallet, etc. The main capabilities, purpose and functions of these information systems are accumulated in the created electronic catalogue, which is one of the main components of the teaching materials. Information systems are selected and placed in the catalogue according to the developed classification. The description of information systems in the catalogue occurs in a structured metadata format in accordance with the Dublin Core Metadata Initiative scheme. A machine-readable form of the catalogue was developed with placement in the information system with support for the Dublin Core Metadata Initiative metadata representation scheme and the OAI-PMH metadata exchange protocol, which makes the catalogue available for indexing by search engines and integrates it into the information scientific space.

The catalogue is designed for use in the research work of graduate students who are faced with the problem of rational and effective selection of scientific information necessary for conducting research work in the preparation of a master's thesis, as well as for the formation of competencies associated with the development of analytical skills in the selection and description of information systems for effective use in solving professionally significant tasks.

3. Conclusion
The implementation of the project will improve the curriculum of the master's program “Digital Technologies of a Smart City”, increase the analytical and instrumental competencies of teachers and staff who carry out scientific supervision of undergraduates and graduate students. The teaching materials will be used in the research work of undergraduates and postgraduates of graduate programs at the Institute of Design and Urban Studies ITMO. Some of its elements will be introduced into educational programs for training bachelors and masters of St. Petersburg State University. The university and the university community will receive a modern universal tool for performing research work, which was previously presented fragmentarily in the educational environment. The sustainability and viability of the project results lie in the versatility of the approach, its independence from software tools, which has been empirically proven by the authors and recognized by the scientific community. The introduction of teaching materials in the educational process began in the approbation phase in September 2020 as part of the discipline “Information technologies in scientific activity”. The EMC will
be fully implemented in the framework of the new academic discipline “Data extraction and mining technologies in scientific research” for graduate students recruiting in 2021.

Acknowledgments
The study was carried out with the financial support of the Vladimir Potanin Charitable Foundation within the framework of project GK200000654.

References
[1] Kononova O V, Krutko E A and Lyapin S Kh 2016 Knowledge Extraction Technologies in the Service of Research Activities at the University Information Society 6 25-37
[2] Kononova O and Liapin S 2016 Using the Contextual Search for the Organization Scientific Research Activities Communications in Computer and Information Science 674 392-9
[3] Kononova O and Prokudin D 2018 Synthetic Method in Interdisciplinary Terminological Landscape Research of Digital Economy Proc. of SHS Web Conf. 50 01082
[4] The fourth industrial revolution. Popular about the main technological trend of the XXI century 2017 TAdviser http://www.tadviser.ru/index.php/Article:Fourth_industrial_revolution_(Industry_Industry_4.0)
[5] Hermann M, Pentek T and Otto B 2016 Design principles for industry 4.0 scenarios Proc. of 49th Hawaii International Conference on System Sciences (HICSS) https://www.researchgate.net/publication/307864150_Design_Principles_for_Industrie_4.0_Scenarios_A_Literature_Review
[6] Schwab K 2017 The Fourth Industrial Revolution (New York, NY: Crown Business)
[7] Dragicevic N, Ullrich A, Tsui E and Gronau N 2019 A conceptual model of knowledge dynamics in the industry 4.0 smart grid scenario Knowledge Management Research & Practice
[8] Massimo Melucci Contextual Search: A Computational Framework, Foundations and Trends in Information Retrieval vol 6 4-5 257-405
[9] Elkina E E, Kononova O V and Prokudin D E 2019 Typology of contexts and principles of contextual approach in interdisciplinary research Modern Information Technologies and IT Education 15(1) 141-53
[10] Kononova O V, Prokudin D E and Elkina E E 2019 Contextual knowledge extraction: terminological landscape of digital economy Sociology of Science and Technology 10(3) 103-20
[11] Kononova O V and Prokudin D E 2020 An approach to the extraction, explication and presentation of contextual knowledge in the study of developing interdisciplinary research areas International Journal of Open Information Technologies 8(1) 90-101
[12] Romanovna T I, Sharapovna K N, Nikolaevna R M and Yurievna I O 2020 The digital economy and the digitization of – past, present, and future International Journal of Psychosocial Rehabilitation 24(8) 7492-507
[13] Grodskiy V S and Khasaev G R 2020 Digital economy – information era: Retrospective analysis Advances in Intelligent Systems and Computing 908 163-79
[14] Kolesnikov A V, Zernova L E, Degtyareva V V, Panko I V and Sigidov Y I 2020 Global trends of the digital economy development Opcion 36(26) 523-40