

Synthetic Method in Interdisciplinary Terminological Landscape Research of Digital Economy

Olga Kononova^{1,*}, and *Dmitriy Prokudin*^{2,1}

¹ ITMO University, Russia

² Saint Petersburg State University, Russia

Abstract. The science in the digital age focuses on interdisciplinary research as the main driver of the development. Modern studies of semantic and thematic contexts taken from the digital scientific resources are important for tasks of information technology use in science, education, management and business. As a rule, interdisciplinary research emerges in response to the trends in social and economic development. Socio-political and cultural environment has also a great influence on the introduction of new terminology into the scientific discourse. The delay of scientific discourse regarding the public discourse generates problems of both scientific and philosophical nature. Furthermore, those challenges influence the education system, making demands for new curricula and courses, which, among other things, reflect the categorical and conceptual basis of the new subject areas. The actual objective is to reveal the trends in the terminological landscape of developing interdisciplinary research directions. The present study offers the Synthetic Method, which combines various approaches and tools of Digital Humanities. This Method allows using context query for the search for information resources; the context knowledge explication and clarification of the terminological landscape of interdisciplinary scientific direction "Digital Economy: e-governance and smart technology".

1 Introduction

The generation of huge arrays of semi-structured and unstructured information that shows both changes in social development and reflections of these changes by the scientific community characterizes the present global information society. The goal of modern interdisciplinary studies is analysis of perspective scientific trends, forecast demand for their results in various areas of public life. One of the main trends of modern interdisciplinary studies is explication of contextual knowledge through the application of methods, approaches, technologies and tools of Digital Humanities.

Technologies and tools in interdisciplinary studies are determined by the broad development of network and distributed access to the information resources (scientific data and knowledge), and high speed of knowledge up-to-date. These days we can observe lagging behind the development of science-metric disciplines from the growth of interdisciplinary scientific terminological corpus, which is generated unmanageably by research schools, groups and individual researchers. Unstructured information, even with free access, and polysemy of the used terminology make it impossible for researchers to track emerging trends and relationships efficiently. This leads to the loss of the important part of scientific knowledge and hypotheses that have not yet received

immediate and wide spreading. Therefore, the study of the terminological landscape development of interdisciplinary scientific areas is relevant and timely.

There is a little research of the development of the categorical conceptual apparatus in various interdisciplinary scientific areas, for example, "The Change of Definitions in the Multidisciplinary Landscape: the case of human embryo and pre-embryo identification" [1]. As a rule, in such works, methods and approaches of Digital Humanities are used incompletely. The application of the Synthetic Method in similar studies is more effective because that Method involves the use of information and communication technologies (ICT) at all stages of the research: from searching and collecting information to its quantitative and qualitative processing.

The article reveals the possibilities of searching and extracting metadata and texts from digital information resources for further quantitative and qualitative processing using the Synthetic Method. The terms 'Digital Technology' and 'Smart Technology' as the most relevant to the State Program "Digital Economy of the Russian Federation" are chosen as the main ones for the search query.

2 Review of methods, technologies and tools

* Corresponding author: kononolg@yandex.ru

In modern dynamically developing information society, the basic mechanism of information volume increase is moving to its digital representation. The use of various technologies and tools of textual information representation in digital form leads to acceleration processes of generating information and increase in volume of electronically stored information. Information and communication technologies are mostly used to access such information. The development of communication technology has resulted in global and broad access to accumulated data of digital information. It is fully applied to scientific knowledge, which is now represented in digital forms [2-5].

Technologies aimed at scientific knowledge accumulating and providing on-line access to it are the most developed ones. Digital scientific resources generally have a powerful search engine that allows searching the necessary information according to specified criteria quickly and efficiently. Access to data arrays distributed in digital scientific Internet resources provides a solution to the problems of availability and rapid dissemination of research results [6], [7]. Technologies of digital scientific resources give a possibility to solve the problem of long-term storage of scientific information [8-10]. Contextual knowledge (i.e. knowledge contained in various contexts, derived from the full-text queries) can exist in various types and forms; can be extracted and studied as a result of the full-text search and its subsequent processing. In the world, scientific discourse study of contextual knowledge is carried out primarily in two directions: the theory and practice of content analysis; the theory and practice of context learning.

Study of contexts of different types, forms and levels are carried out in the frame-work of the traditional content analysis: the method and technology of qualitative-quantitative analysis of documents in order to identify or measure social facts and trends reflected in those documents. Content analysis examines the documents in the social context. Content analysis of media reports based on the paradigmatic approach is increasingly distributed. According to this approach studied signs of texts (content problems, causes, the problem-formatted subject, the degree of problem tension, the ways of its solution, etc.) are regarded as a definite organized structure [11].

The tasks of the context knowledge study are fulfilled through the approaches described in Table 1.

Moreover, one more important research direction is the analysis of heterogeneous data processing capabilities (texts), distributed in heterogeneous information systems with access from both global and local networks. This field is aimed at implementing Grid technologies information distribution to its systematization and integrity at its constant quantitative augmenting [12, 13].

Table 1. Approaches of the contextual knowledge study.

Approaches	Objectives
Technological	the developing of information systems and implementing contextual search algorithms in search engines [14, 15]

Semantic	the developing and using of linguistic methods to the analysis of tests and identification of the certain meanings in those tests [16, 17]
Content	an applied use of information search and analysis algorithms in functioning information systems for quantitative and qualitative analysis of their content processing from the definite subject domains [18]

These days quite a lot of software products targeted at providing specialized services for handling unstructured textual and non-textual information are worked out. Basically, those are the programs of linguistic analysis of the text. They are designed for special tasks for text analysis, but not all of them provide full text search. Their list and detailed description is presented on <http://asknet.ru/analytics/programms.htm>.

There are a few main projects of contextual knowledge research, implemented in the last 10 years by Russian university, specialized laboratories, and research centre.

1. The Sociological Institute of the Russian Academy of Sciences (Saint Petersburg): Context-oriented Methods to Construct Theories (2006-2008); Context-oriented Ontologies of Sociological Research (2009-2011); Working out Constructive Methods to Develop Sociological Knowledge (2012-2014). Graph context-oriented ontological methods of knowledge management are used to solve problems of operating with sociological concepts and explicit relationships between them by creating relation-ships of computer functionality (inheritance, encapsulation, type-creating, etc.) [24]. Keeping preterition, checking the concepts polysemy, maintaining their identity being used in different contexts; analytical estimating methods of qualitative research thesaurus is supported. The research package of applied computer programs, used in listed research, program-simulating the proposed methods, is not an intuitive clear tool. Therefore, it requires prior preparation. This approach is useful for performing large scale modelling projects of subject domain knowledge.

2. The Higher School of Economics (Saint Petersburg): working out the concept and methodology of multilevel monitoring of the state of interethnic relations according to social networks (2015). Suggested approaches and tools are not intended to identify process and analyse scientific information data, search and explication of contexts respectively the analysing field of knowledge.

3. The Institute of Informatics Systems named after A. P. Ershov of Siberian branch of Russian Academy of Sciences (Novosibirsk). Artificial Intelligence Laboratory: Research and development of methods and tools for analysis and visualization of heterogeneous knowledge of large information portals (2009-2011). The concept of thematic intellectual scientific Internet resource (ISIR) was suggested for informational and analytical support of the scientific and productive activity in a particular field of knowledge. ISIR technology is targeted to professionals in different fields of knowledge. It represents the development of the

previous technology for building portals of scientific knowledge. The structure of the electronic Russian-English Thesaurus on Computer Linguistics is worked out. The thesaurus is a complete and consistent system of concepts from the field of Computer Linguistics, linked by semantic relations, reflecting the position of each notion in this system.

4. The Institute for System Programming named by V. P. Ivannikov, Russian Academy of Sciences (Moscow). The Institute has developed a software system Texterra to use Wikipedia for enhanced search and navigation in text database. Texterra is an effective system from computational point of view. It uses semantic proximity measure for information search and improvement of the query results ranking. Semantic information accumulated in Wikipedia-type databases allows improving methods of automatic text processing and information retrieval. Instead of keywords, domain specific terms are used. Unfortunately, the effectiveness of such search engine is not possible due to the lack of effective methods of indexing documents for query execution. Modern methods of calculating semantic similarity in English are well described in the following studies: [10, 19, 20]. The tool does not support the network distributed environment.

5. University ITMO (Saint Petersburg). T-Libra (Humanitarian project): creation of a virtual information and resource centre for the extraction of knowledge from humanitarian texts based on advanced full-text search, and functional integration of resources and services in a distributed environment (2014-2016) [21, 22]. It supports operating in a network distributed environment and allows carrying out search, explication, integrated analysis of the context knowledge. Moreover, it implements the possibility to build thematic trends of concept trends.

Approaches, models and methods of search for information resources with use of ICT are discussed in publications [12, 13, 16, 21 – 27]. The publication analysis has revealed that comprehensive research, which would integrate all those instruments, has to be held. The considered research offers solutions that presuppose the existence of formalized conceptually structured base and categorical apparatus of interdisciplinary research. The last has to contain thesaurus – set of interrelated keywords or term-concepts ontology relevant to the field of research. Methods and search tools corresponding them, explication and analysis of context knowledge, which have been implemented in individual information systems and web-services, remain unknown and unused due to the lack of meaningful information about the tools as well as the technique and algorithms of their use in science. Low demand of such tools is explained by the lack of thematic study for the selection and systematization of meta-descriptions, thesaurus, subject domain ontologies that form database and knowledge of those tools. The extra reason is implementation of a limited set of methods for searching and interpreting information in each separate released tool.

3 Synthetic method of contextual knowledge research and retrieval

A synthetic method of contextual knowledge extraction is an integrated approach oriented to solve problems of allocation and explication of scientific content on topical directions of interdisciplinary scientific research. The synthetic method uses advanced systems of full-text and multimodal search of the network distributed environment. The method involves the extraction, expert evaluation and interpretation of contextual knowledge from large text and non-text information data. This method makes it possible to enter a new level of understanding the typologies of context knowledge, formation, objectification and updating of terminology basis of interdisciplinary research.

The developed integrated approach, named here Synthetic Method, displays the content analysis to a higher level. In the traditional content analysis, the target function and category analysis are of primary importance, while derived generalized text units of analysis are more important. For the Synthetic Method, a generalized text (with elements of multimodal information) has a priority, the resulting content – structured description of contextual knowledge is secondary. It can be said that the traditional content analysis and proposed synthetic analysis are complementary methods and technologies of learning content and semantic information contexts. Generic text (text and multimodal information) in this case is a generator of explicated contexts and con-text knowledge structures. The method being developed contains:

- a primarily explication of the corpus (expertise of the terminological base: dictionary, tutorials, manuals, program documents);
- a construction of mind maps;
- a method of representative samples selection of scientific texts of different origin and location from open sources;
- a step-by-step instruction of using combinations of well-known search methods, technologies and tools of automated extraction and study of contextual knowledge;
- an expert assessment algorithm of selected paragraph- contexts, which is original and simple for users, enables researchers to generate personal thematic collections of materials;
- construction technology of thesauri of interdisciplinary scientific direction and thematic trends within those fields;
- building the corpus ontology and creation of thesauri of interdisciplinary scientific direction;
- a mechanism of integration of the results in selected environment, supported by recommendations on the interpretation and use.

Unlike existing developments, expected results of applying the synthetic method are supposed to examine the structures and functions of knowledge in interdisciplinary research deeper as well as to use the contextual framework for interpreting scientific texts,

algorithm development, software intelligent search and artificial intelligence.

4 The Digital Economy subject domain

4.1 The tasks of the research

The study objectives are to explore the explication possibilities of scientific knowledge gained from the digital information resources of media and scientific resources bases, and then to extract main contexts of the research subject domain of the interdisciplinary scientific direction "Digital Economy: e-governance and smart-technology". According to the synthetic method, the primarily tasks should be solved:

- Justification of applied environment selection to ensure the objectives of the study in accordance with the established model of contextual knowledge explication.
- Forming the semantic core (key term-concepts) of the subject domain.
- Study of the possibility of using digital information resources for forming arrays of Russian scientific texts which are relevant to subject domain (publishing platforms: eLibrary, Cyberleninka, Socionet, Integrum, EBS "Lan", East View); characteristics of resources, criteria influencing the choice, politeness [28].
- Study of digital Media information resources.
- Comparative analysis of the terminological landscape and dynamics of term-concepts' using of Media and scientific resource bases.
- The main context explication and the formation of relevant full-text arrays for further analysis, construction of trends, structure and meta-description of thesaurus of the subject domain.

4.2 Terminological core of the domain

The preliminary stage of the current research is the selection of the terminological core of the "Digital Economy" subject domain. The analysis of the State Program "Digital Economy of the Russian Federation" (approved by the Decree of the Russian Federation Government dated July 28, 2017 № 1632-p) was carried out to initially identify the terminological landscape of the interdisciplinary direction "The Digital Economy: e-governance and smart technology." The analysis results are shown in Table 2.

The main conclusion is the existence of independent term-concepts 'Digital Economy', 'Digital Technology' and 'Smart Technology' as integral parts within the framework of the Digital Economy subject domain. Therefore, the research focuses on both searching for full-texts, which use key term-concepts for the subsequent detection and analysis of the contexts of their use, and determination of dynamics of the interest growth and the spectrum of interest to the Digital Economy subject domain in the scientific community and society.

Table 2. Terminological core of the State Program "Digital Economy of the Russian Federation".

Term-concept	Subject domain					
	Computer science	Economics	Sociology	Education	Political science	Management
Citizen			+		+	+
City (Smart)	+					+
Data (Big, Personal)	+		+			+
Ecosystem	+	+	+			+
Services (State, Municipal)		+				+
Society			+			+
System (Information)	+					+
Technology (IT, ICT)	+		+	+	+	+
Technology (Smart)	+					+
Technology (Digital)	+					+

Analysis of the submitted terminological core reveals that Digital Economy continues the development of ideas and projects, which have appeared in Russia since 2005 and can be characterized by term-concepts: 'Smart City', 'State Services' (including 'Municipal Services') 'Ecosystem', 'Big Data', 'Citizen' and 'Society'. In the "Digital Economy" interdisciplinary direction the term-concepts 'Information Technology', 'Info-Communication Technologies' are more frequently used. Therefore, the directions also include such term-concepts as 'Smart Technology' and 'Digital Technology'. All term-concepts were also evolved in the context of the strategy, approved by the Decree of the President of the Russian Federation in May 9, 2017 № 203 "About the development strategy of information society in Russian Federation at 2017-2030".

4.3 The analysis of the array of digital Russian resources

For the further study of the development of the interdisciplinary direction corpus, the authors assessed the relevance of the main terms in Russian full-text digital electronic resources to the basic concepts. Those resources include: scientific publications (eLibrary, Kiber-Leninka, Socionet, East View); scientific, popular scientific and educational-methodical literature (EBS Lan, East View); as well as Media – newspapers and magazines (Integrum, East View). Using the built-in search tools, we get the following results (Table 3).

Analysis of the results from eLibrary and East View indicates that in both scientific and socio-political discourse, the considering term-concepts appeared in the early 2000s that corresponds to the beginning of the information society development in Russia as a sustainable social trend.

To evaluate the dynamics of the terminological landscape of the reporting interdisciplinary direction, search for scientific publications on the term-concepts of

‘Digital Economy’, ‘Digital Technology’, and ‘Smart Technology’ has been carried out in the scientific electronic library (eLibrary). The sample chosen includes publications since 2005 to 2017. The results obtained are spread across the dates as shown in Figures 1-2.

Table 3. Distribution of publications on major term-concepts in Russian full-text resources.

Digital resources	Term-concepts ^a , frequency of use			
	DE	DE& SS	SS& IT	SS& DT
eLibrary	1124	6	826	44
Cyberleninka	606	66	2942	11
Socionet	80	2	28	0
Integrum	5121	132	14398	436
EBS “Lan”	1057	312	1085	580
East View	141	5	151	8

A search was made for Russian-language digital resources of two types: media on the Integrum resource platform (reflects the socio-political discourse) and scientific publications on the eLibrary resource platform. Search queries were made on term-concepts ‘smart technology’ and ‘digital technology.’ At the same time, in eLibrary, the search was carried out according to thematic headings "Economics", "Computer science", "Computer Engineering", "Organization and management". This made it possible to select the publications most relevant to the direction under consideration.

The Integrum resource resulted in 1282 documents from 3534 publications. To extract metadata and full texts, publications with at least 5 publications were selected. The maximum number of publications (25) was published in the newspaper “Komsomolskaya Pravda” and regional economic digest "Region-Inform." 15 editions were selected with a total of 227 publications. The upload was made in a text file of *.rtf format for further contextual analysis.

Dynamics of publications shows a steady growth in the use of the all term-concepts. The term-concept ‘Digital Economy’ has started entering scientific discourse since 2010. The peak of its use is accounted was in 2017. The cover of the use of the term-concept ‘Digital Technology’ grows similarly (Figure 1).

To study the development of the terminological landscape of the interdisciplinary direction under consideration, the term-concepts 'Digital Technology' and 'Smart Technology' of Russian-language full-text digital electronic resources are relevant. The largest publications are scientific (eLibrary) and mass media such as newspapers and magazines (Integrum). The search was carried out from 2005 to 2017 (Figure 2).

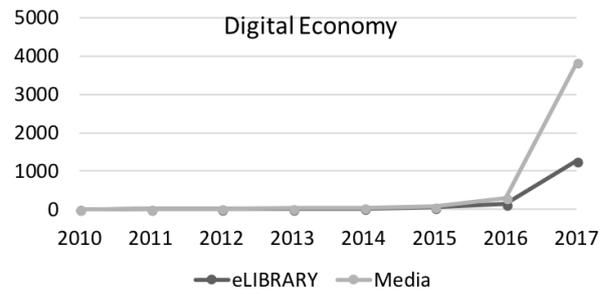


Fig. 1. Dynamics of publication quantity (term-concept ‘Digital Economy’) in electronic resources of eLibrary and Media.

For clarity, graphics for the term-concept 'Digital Technology' have been constructed on the minor axis (secondary axis, on the right), since the total number of publications with that term-concept in the Media is much larger than the rest of the analyzed data.

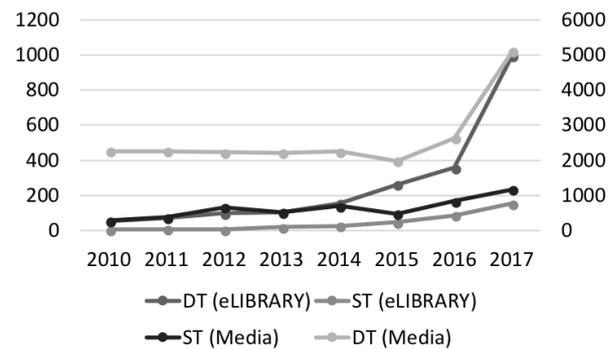


Fig. 2. Dynamics of publication quantity (term-concepts ‘Digital Technology’, and ‘Smart Technology’) in electronic resource of eLibrary and Media.

Analysis of the data obtained has shown that until 2016, the mention of terminological concepts is of even character, i.e. interest in these subjects is stable. A sharp jump begins in 2016, when the government initiated the State Program "Digital Economy of the Russian Federation". There is an increase in the discussion of ‘Smart Technology’ both in social and political discourse, and in the scientific one. The same sharp surge is also characterized by the use of the term ‘Digital Technology’, which indicates an increased interest in them in the context of the development of the Digital Economy domain.

4.4 Clarifying the terminological context of the domain core

For initial clarification of the semantic kernel of the term-concepts in the certain interdisciplinary domain, a pilot study has been held where the Russian information-analytical web portal eLibrary (<http://elibrary.ru>) is used. Information arrays are obtained by bringing the keywords 'Digital Economy', ‘Digital Technology’, and ‘Smart Technology’ into the search bar. The advanced search is limited by the search in Russian scientific

^a 'Digital Economy' - DE; 'State Services' – SS; 'Information Technology' – IT; 'Digital Technology' - DT.

journals, conference proceedings and scientific thesis as well as by the dates of publication from 2005 to 2017.

The interest in dynamics of those term-concepts' using is associated with the emergence of socio-political and scientific discourse around government initiatives to develop Digital Economy in Russia (Tables 4 – 6).

Table 4. Term-concept 'Digital technology'.

Indicators		Digital technology		
		2011	2014	2017
Total number of publications		73	157	995
Number of Publications in journals	indexed in WoS or Scopus	3	8	29
	referred to the RICC core	10	22	70
Thematic rubrics	Economy. Economic sciences	13	44	546
	Education	4	15	69
	Computer Engineering	5	19	54
	Computer science	5	10	48
	Organization and management	4	-	19
Key words	Digital Economy	-	-	272
	Digital Technology	10	16	154
	IT, ICT	0	18	62
	Blockchain	0	0	44
	Innovation	3	-	35
	Information society	3	-	-

Table 5. Term-concept 'Smart technology'.

Indicators		Smart technology		
		2011	2014	2017
Total number of publications		9	26	149
Number of Publications in journals	indexed in WoS or Scopus	2	0	3
	referred to the RICC core	1	1	9
Thematic rubrics	Economy. Economic sciences	5	9	83
	Education	2	-	6
	Computer Engineering	2	4	10
	Computer science	-	2	4
	Organization and management	-	-	4
Key words	Digital Economy	-	-	14
	Digital Technology	10	16	154
	IT, ICT	-	2	7
	Blockchain	-	-	-
	Innovation	2	4	8

Table 6. The dynamics of the publish activity and the publication' status.

Indicators	Smart technology			Digital technology		
	2011	2014	2017	2011	2014	2017
Total number	9	26	149	73	157	995
Indexed in WoS, Scopus	25%	0%	3%	4%	6%	4%
Included in the RICC core	13%	5%	9%	15%	16%	10%

Analysis of scientific resource of eLibrary for more than 10 years has shown the steady growth of the publication quantity, which has used selected key words. There were very few publications earlier than 2011 and none in the period of 2005-2006.

Search queries (from 2011 to 2017) made on publication array about technologies of the Digital Economy allows identifying 10 sustainable contexts: 'Blockchain', 'Digital Economy', 'Digital Technology', 'Innovation', 'Information Society', 'Internet', 'Internet of Things', 'Smart Technology', 'Smart Society', 'Smart City'. The conducted experiment results suggest that the description of the terminological landscape of the interdisciplinary direction "The Digital Economy: e-governance and smart technology" besides the selected term-concepts is necessary to be supplemented by the following ones: 'State Information System', 'e-Government', 'State Services' as well as 'Big Data.'

Everyone from them should be considered as a complex object. Further research suggests the full-text analysis of the publication arrays within the framework of Synthetic Method. In this regard, for the extraction and contextual analysis of Russian full-text bases the efforts of the researchers are concentrated on the affordable distributed network environment of ITMO University.

Conclusions

The main results of the conducted research with use of Synthetic Method for context knowledge studying on the example of the text array of interdisciplinary scientific direction "Digital Economy: e-governance and smart technology" are the following ones:

- There has been revealed the possibility of the extraction of digital information resources for replenishment of distributed information scientific environment T-Libra in the researched subject domain. The main sources are eLibrary (scientific publications), Integrum (Media) and East View (scientific publications and Media).

- The mechanisms of terminological landscape formation and extension of the semantic core (conceptual base) of the subject domain have been elicited. Conceptual base has been expanded by the new term-concepts.

- It has been shown the possibility to study the dynamics of digital information resources' using with the tools of eLibrary.

The practical significance of the results is that they can be used to develop and modernize educational programs, training specialists, who will participate in the implementation of the State Program "Digital Economy of the Russian Federation" in practice, as well as to develop and refine actual thesauruses of that modern subject domain.

Further studies can pursue the objectives in order to expand the resource base they are held on and apply the

appropriate context search services in the distributed information scientific environment T-Libra.

The reported study was funded by RFBR according to the research project "The integrated approach elaboration to the analysis of the terminological base of the developing interdisciplinary research in distributed network environment" № 18-011-00923, 2018-2020.

References

1. C. Piciocchi, L. Martinelli, *Croatian Medical Journal* **57(5)**, 510 (2016)
2. C. L. Borgman, *Digital Humanities Quarterly* **3(4)** (2009)
3. *Digital Libraries and Information Access: Research Perspectives* (Neal-Schuman/ALA, Chicago, Illinois, 2012)
4. M. Weller, *The Digital Scholar: How Technology Is Transforming Scholarly Practice* (Bloomsbury Academic, London, 2011)
5. M. Laakso, *Scientometrics* **99(2)**, 475 (2014)
6. E. Scanlon, *Br J Educ Technol* **45**, 12 (2014)
7. M. Agosti, L. Benfante, N. Orio, *Lecture Notes in Computer Science* **2911**, 253 (2003)
8. D. Burda, F. Teuteberg, *Journal of Information Science* **39(4)**, 442 (2013). DOI: 10.1177/0165551513480107
9. H. J. Nielsen, B. Hjørland, *Journal of Documentation* **70(2)**, 221 (2014)
10. T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, *Proc. Advances in Neural Information Processing Systems* **26**, 3111 (2013)
11. O.T. Manaev, *Content analysis as a research method*. <http://psyfactor.org/lib/content-analysis3.htm>, last accessed 2018/02/21.
12. L.I. Borodkin, *Krug idej: mezhdisciplinarnye podhody v istoricheskoj informatike, Trudy X konferencii Asociacii "Istorija i komp'juter"* (Moscow, 2008)
13. O.L. Zhizhimov, Y.I. Molorodov, I.A. Pestunov, V.V. Smirnov, A.M. Fedotov, *Vestnik NSU. Series: Information Technologies* **9(1)**, 67 (2011)
14. G.S. Osipov, I.A. Tihomirov, I.V. Smirnov, *Programmnye sistemy: teorija i prilozhenija. Trudy mezhdunarodnoj konferencii. Programmnye sistemy: teorija i pri-lozhenija. IPS RAN, Pereslavl'-Zalesskij, 1, 21* (Fizmatlit, Moscow, 2004).
15. F. Tao, K.L. Hou, J. Han, C. Zhai, X. Cheng, M. Danilevsky, N. Desai, B. Ding, J. Ge Ge, H. Ji, R. Kanade, A. Kao, Q. Li, Y. Li, C. Lin, J. Liu, N. Oza, A. Srivastava, R. Tjoelker, C. Wang, D. Zhang, B. Zhao, *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '13)* (ACM, New York, NY, USA, 1494, 2013)
16. A.V. Chernij, A.F. Tuzovskij, *Znanija-Ontologii-Teorija* **2**, 52 (Novosibirsk, ZAO "RIC Prajs-Kur'er", 2009)
17. P.D. Turney, P. Pantel, et al., *Journal of Artificial Intelligence Research* **37(1)**, 141 (2010)
18. H. Saifa, Y. Heb, M. Fernandez, H. Alani, *Information Processing & Management* **52(1)**, 5 (2016)
19. E. Agirre, E. Alfonseca, K. Hall, J. Kravalova, M. Paşca, A. Soroa, *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL '09)* (Association for Computational Linguistics, Stroudsburg, PA, USA, 19, 2009)
20. E. Bruni, N. K. Tran, M. Baroni, *Journal of Artificial Intelligence Research* **49**, 1 (2014)
21. S. Kh. Lyapin, A.V. Kukovyakin, *Collection of scientific papers XVIII of the scientific conference "Internet and Modern Society"*, 216 (ITMO University, St-Petersburg, 2015)
22. O. Kononova, S. Lyapin, *Communications in Computer and Information Science* **674**, 392 (2016)
23. Z. N. Apanovich, A. M. Marchuk, *HIV National Conference on Artificial Intelligence with International Participation KII-2014 (October 24-27, 2014, Kazan, Russia): Proceedings of the Conference (RIC «Shkola», Kazan, 1, 2014)*
24. A. V. Buhanovski., V. N. Vasilyev, *News of higher educational institutions. Instrument making* **53(3)**, 60 (2010)
25. G. V. Kanygin, M. S. Poltinnikova, *Trudy SPIIRAN* **48**, 107 (2016)
26. W. P. Taylor, *A comparative study on ontology generation and text clustering using VSM, LSI, and document ontology models* (Clemson University, 2007)
27. Y.A. Zagorulko, O.I. Borovikova, *Vserossijskaja konferencija s mezhdunarodnym uchastiem "Znanija – Ontologii – Teorii" (ZONT-2011)* (Novosibirsk, 2011)
28. D. Prokudin, G. Levit, U. Hossfeld, *Internet and Modern Society: Proceedings of the International Conference IMS-2017 (St. Petersburg; Russian Federation, 21-24 June 2017)* (ACM International Conference Proceeding Series, ACM Press, N.Y., 2017)