Author's Choice for Keyword List: Research Aspect *

Olga Kamshilova[1]
onkamshilova@gmail.com

Larisa Beliaeva[1]
lauranbel@gmail.com

Lyubov Geikhman[2]
lyubageykhman7@gmail.com

[1] Herzen State Pedagogical University of Russia, Saint Petersburg,
[2] Perm National Research Polytechnical University (PNRPU), Perm, Russian Federation

Abstract

The paper addresses the problem of creating a relevant keyword list prefixing a research article. It discusses the issue through the lenses of informational, psycholinguistic and editorial concepts of keywords. It considers keywords as a text form within the modeled text of a research paper and deals with their information value. Based on a quantitative text analysis of current publications presented by Russian authors (8 case studies), the research shows at least three strategies of compiling a keyword list of which none is absolutely successful or meets the demands of publication promotion. Candidates for key words are multidimensionally treated in respect of a) their relation to other compositional parts of research text, b) their morphological and syntactic character, c) their information value according to query statistics.

Keywords: *keyword list, keyword set pattern, keyword density, search engine statistics, document search image, text informational image*

1 Introduction

Keyword is an interdisciplinary term with more or less common meaning of "a word that tells you about the main idea or subject of something" [Oxford Dictionary]. More common nowadays is the meaning of "a word or phrase that you type on a computer keyboard to give an instruction or to search for information about something" [ibid.]. The problem of keywords is that of their almost routine presence in general discourse: we, at once, seem to know a lot about keywords, enough to understand their prominent text value, but have very little practical knowledge when it comes to defining a set of keywords for our own text, e.g. for a paper or research article. Unlike other parts of a research paper that are prescribed by journal formats, keywords are only regulated by their number limit, which definitely backs up our claim that authors must be well aware about keywords choice proper.

Academic use of the term has a long history (hence, its different spelling interpretations: key word, key-word, keyword) and is associated with at least four research paradigms: cross-linguistic studies [Wierzbicka 1997, Williams 1983, Stubbs 1996, Stubbs 2001, Zemskaya 1996, Stepanov 1997, Shmelev 2002], style and text interpretation [Arnold 1999; Lukin 1999; Bolotnova 2004], psycholinguistic investigations of child speech and speech impairment [Sakharny et al. 1984; Sakharny, Stern 1988, Murzin, Stern 1991], and computer and information studies (information retrieval, automatic text processing and abstracting, etc.) [Busa 1980; Ripp, Falke 2018]. We claim to trace these interpretations in authors' actual choice for the keywords list they prefix their conference papers or journal articles.

The linguo-cultural and cross-cultural approach develops the idea that languages are sensitive indices to the cultures they belong and every language has key concepts, which reflect the culture core values. These key concepts are expressed in key words (sic!). Thus, different cultures possess a "natural semantic metalanguage" [Wierzbicka 1997], that helps to study, compare and explain cultural identity and cultural differences. With appeal to anthropologists, psychologists, and philosophers, as well as linguists, a coherent theoretical framework based on multilanguage empirical evidence was introduced for cultural patterns studies.

Similar studies were done as early as 1935 with sociolinguistic focus on "sociologically important words, what one might call *focal* of *pivotal* words" [Firth 1957, 10] (see also [Williams, 1983]). Key words in stylistic and interpretative text studies (I.V. Arnold and others) are treated as topical lexemes, often grouping around one or several lexemes that mark the author's artistic message and help to understand the text informational structure and appreciate its artistic value.

Since we here focus on keywords informational function, it is reasonable to refer to their computational (informational) and linguistic characteristics. The linguistic (originally psycholinguistic) understanding of keywords by L.V. Sakharny and his colleagues was much inspired by the 1970s-1980s progress in information retrieval studies [Baeza-Yates R., Ribeiro-Neto 2011, Manning, Raghavan, Schutze 2008]. The idea of keywords (taken as a set) shaping the so called "document search image" implies that keywords bare the essential information about the text and it is taken as basic for coordinate indexing method [Lindemann, Kliche, Heid 2018; Zeni et al. 2007]. This method claims that a text content may be presented by a list of keywords reflecting its topic with a guaranteed sufficient degree of accuracy and completeness.

In psycholinguistic studies the idea of a keyword set representing the whole text was first transformed into the idea of "a primitive text" or, rather "text-primitive" [Sakharny et al.,

1984], characteristic of early child speech. Later in [Sakharny, Stern 1988] it was resumed as a text category in linguistic description of "synonymous texts" or texts-synonyms that present a dynamic paradigm "title – keywords – abstract – text body". It was then that the authors pinpointed keyword set characteristics that we find relevant for the present study: it was proved that a set of keywords has essential text characteristics that provide it with integrity and cohesion, namely: it has a syntagmatic character (non-random, linear, a sort of "chain" structure) and thematic progression (the keywords mark the text topic (theme) and their order – through associative connections – rheme).

However, understanding keywords as document search image is prevailing in today's practice: "A keyword is a word in a text that, with other keywords, can represent the text <...> A set of document keywords is called a document search image. The set of keywords is close to annotation, plan and abstract, which also represent a document with less detail, but lacking a syntactic structure"[1]. To assist information retrieval from big data keywords may be obtained by linguistic and computational methods (e.g., by analyzing word frequency in the text). In personal practices keywords are simply a query that a person prints into a search engine window when they want to find something, or special HTML tags that we may add to our texts (blogs, sites, chats, MS Word texts, etc.), which search engines, generally focusing on keywords highlighted as a result of their analysis, shall take as additional information.

The above-mentioned approaches reveal keywords essential characteristics that they demonstrate by their discourse function, so it may be reasonably suggested that a conscious author must have a most general idea of what a keyword is. But, evidently, neither everyday search practice, nor journal or conference paper regulations are of much help when authors decide on their choice of candidates for the keyword list to prefix their texts meant for publication. In attempt to find out whether there are some patterns behind this choice we try to answer the following questions.

1. What candidates do authors choose for the keyword list?
2. How do the keywords show in the text and what is their proportion in the text body?
3. How do the keywords promote the publication informationally?
4. What's the actual need for keywords if any at all?

2 Data and Methods

Data used in this study are 8 sample Russian texts of research articles on different subject areas in humanities submitted to a reviewed university journal. The rationale for concentrating on a comparatively small number of texts in this paper is that the suggested analysis implies taking text as a major analysis unit, including minor ones, such as title, keyword list, abstract and keywords proper. Thus, every text in the selection provides a basis for a case study.

We consider a research paper to be a modeled text, written according to a conventional structure, though conventions in Russian academic tradition are not as developed as AIMRaD pattern, and following [Sakharny, Stern 1988] we take *title, abstract, keyword list* and *text body* as texts-synonyms representing a dynamic paradigm.

The first step was to provide *a quantitative description of keywords presence* in the text. For this purpose, we attempted to state the number of keyword occurrences in the text body,

[1] http:/www.seobuildung.ru

as well as in all synonymous texts within the whole text of the research paper. The search of texts and word lists (frequency), concordance plot analysis for every text and its units under discussion were performed with AntConc toolkit that provides reliable results for Cyrillic texts.

On the second step we analyzed the *keyword list as a set*, *a primitive text* with its possible linguistic characteristics – syntagmatic and thematic relations – in order to trace patterns or authors' preferences in compiling the keyword list.

Thirdly, we defined keyword density for some of the keywords in every text so as to judge their informative value [Passonneau 2006]. Keyword density is the percentage of times a keyword appears in the text compared to the total number of words in the text. In the context of search engine optimization, keyword density is used to determine whether a web page is relevant to a specified keyword or keyword phrase. In our context it is a way to establish the information value of authors' keywords, their ability to assist the text promotion on the Internet.

3 Results

3.1 Keywords presence in the text

Treating the keyword set as a text within a dynamic paradigm of 4 synonymous texts one can expect repetition of the key words in every part of this paradigm. Searching authors' keywords in the texts and their parts soon gave evidence enough that keywords in the 8 texts demonstrate various degree of their presence in the texts, as well as a quite unspecific distribution among the text parts (synonymous texts). In some texts keywords are missing in the abstract, e.g. in Text 3 and 4, or in the title and abstract, as in Text 8, Texts 5 and 8 demonstrate minimum use of keywords in the whole paradigm (Table 1).

Table 1: Keywords absolute frequency in synonymous texts

	Keyword List	**Keywords in the Title**	**Keywords in the Abstract**	**Keywords in the Text Body**
Text 1	7	5	10	172
Text 2	6	2	4	124
Text 3	5	**1**	**0**	31
Text 4	5	**1**	**0**	23
Text 5	3	**1**	3	**3**
Text 6	3	**1**	2	100
Text 7	5	2	6	40
Text 8	3	**0**	**0**	6

Wordlists analysis alongside with stem search revealed that besides different wordforms of the keywords "keyness" is maintained through the text by:

- derivation (stemming): *иероглифика → иероглифический, иероглифы (hieroglyphics → hieroglyphic, hieroglyphs)*;
- word composition / collocation: *власть → властеотношения (power → power relations)*;

- generalization / specification: *кредитные институты/учреждения – банковско-кредитные учреждения (banking credit institutions → credit institutions).*

Sometimes the keyword from the list is used in the text only in its derivative form (Table 2):

Table 2: Keywords presented in the text by derivatives alone

	Keywords List	**Abstract, Text Body**
Text 3	иероглифика (hieroglyphics)	иероглифический, иероглифы (hieroglyphic, hieroglyphs)
Text 4	банковские кредитные институты (banking credit institutions)	кредитные институты (credit institutions)
Text 7	проблемы международной миграции (problems of international migration)	международная миграция, миграция, проблемы миграции (international migration, migration, problem)

3.2 Keyword sets as a pattern

The composition of elements within the keyword set demonstrates what presumably may be taken for the authors' different strategies in compiling the keyword list. We dare speak of at least three patterns underlying the analyzed sample texts.

One may be traced back to the ideas expressed in [Sakharny, Stern 1988], since the sets of the kind are characterized by linear syntagmatic relations (a sort of "chain" structure) of terms (usually of different but crossing subject areas), very close to text-primitives, moving from more general to specific notions or *vice versa*, as in Text 5. The effect of "telling a story" is achieved by the keywords order that through associative connections reflects the text thematic progression:

- *язык, концепт, эмоция, поле, доминанта, лезгинский язык, английский язык / language, concept, emotion, field, dominant, Lezgi, English* (Text 1);
- *власть, эстетика, красота, возвышенное, трагическое, комическое / power, aesthetics, beauty, sublime, tragic, comic)* (Text 2);
- *китайский язык, машинный перевод, иероглифика, лексическая структура, статистический анализ / Chinese, machine translation, hieroglyphic, lexical structure, statistic analysis* (Text 3);
- *рациональность, наука, знание / rationality, science, knowledge* (Text 5);
- *власть, коммуникация, гражданское общество / power, communication, civil society* (Text 6).

The second is nothing but classification indexing / subject indexing within a definite subject area, which is common for Dewey Decimal Classification (DDC) or consequent Universal

Decimal Classification (УДК), library indexing, library and information science [Bates, Maack 2010]:

- *банковско-кредитные учреждения, кредитный кооператив, кредитное товарищество, ссудо-сберегательная касса, ростовщичество bank-credit institutions / banking and credit institutions), credit cooperatives, credit partnerships, loan and savings bank, usury* (Text 4);

- *профессиональное обучение, профессиональные знания, профессиональные и личностные качества / professional education, professional knowledges, professional and personal qualities* (Text 8).

The third pattern was found only in Text 7, its different composition may be regarded as individual or rather inexperienced, since the position of keywords is presented by the key questions of the article, with morphological (number, case) and syntactic (connection) marking ad hoc: *соотношение глобализации и национализма, проблемы международной миграции, роли национальных государств / correlation of globalization and nationalism, problems of international migration, the role of nation states* (Text 7).

As our observation shows, the linear composition of a text-primitive seems to be preferable for the sample texts authors. As long as Russian academic tradition does not prescribe any regulations for keyword list, we can only suggest that this preference may be explained by natural speech habits, that is by psycholinguistic basis described in [Sakharny, Stern 1988]. The subject-oriented pattern found in Texts 4 and 8 demonstrate a rational strategy that will be addressed to later in Discussion section. The pattern found in text 7 is, presumably, very individual and hardly possible to be published unedited.

3.3 Keyword list as a text information image

Editorial demand for a keyword list to prefix a research publication has an unquestionable informational purpose: it is a guide for the reader to decide on whether the text is of any professional or other interest to him, a reason for the editor to accept and publish the text under the appropriate subject heading, and, in case they match a reader's query, a good chance for the search engine to place the text in the search engine output. In this last case, similar to web sites or other internet issues, to be placed on the first few pages of search result the keywords must be repeated in the whole text often enough, so that their density is not less than 1.5%. According to SEO Keyword Density Analyzer "The optimal density of keywords (and phrases) is from 1.5 to 7%, preferably not more than 3.5%. And at least 2 exact occurrences of the search phrase on the text page"[2].

Obviously, no author of a research article is either conscious of it or previously instructed. The keyword ratio (keyword density) of the most frequent keyword (in case of Text 1 there are two keywords that have almost equal high frequency) for each of the 8 sample texts show that only 3 of them pretend to be hopefully placed on the first pages of search results be the query including them (Table 3).

Even with this relatively high density the "happy" keywords will not bring the text to the top of the search results because the authors' choice of the keyword itself is unhappy, ineffectual: *язык (language), эмоция (emotion), власть (power)* are all very frequent terms

[2]http://site-submit.com.ua/?pg=servis_analizing

of general use that alone (other keywords are of insignificant density) do not specify the text topic, as well as *рациональность (rationality)* does not specify its status of a philosophy category in Text 6. So, none of the 8 texts under study have a keyword list that can be treated as a text search image.

Table 3: Keyword density in the text

	Most Frequent Keyword	Absolute Frequency	Total Number of Words	Keyword Density
Text 1	язык (language) эмоция (emotion)	33 34	1632	**2,02**% **2,08**%
Text 2	власть (power)	66	4441	**1,49**%
Text 3	китайский язык (Chinese language)	17	3662	0,46%
Text 4	кредитный кооператив (credit cooperative)	12	1930	0,62%
Text 5	коммуникация (communication)	9	2526	0,36%
Text 6	рациональность (rationality)	88	4023	**2,19**%
Text 7	глобализация (globalization)	26	1971	1,32%
Text 8	профессиональное образование (professional education)	4	2035	0,2%

There is one more factor that might improve the choice of keyword candidates, namely accounting for the candidate search statistics, that is, how popular queries with the candidate are. Thus, the choice of *иероглифы / hieroglyphs* as a candidate for the keywords in text 3 might have been more effective than the authors preference for *иероглифика / hieroglyphics* (Table 4). Search engine statistics provides the number of queries with the keyword either realized or supposed within a definite period, so it is feasible to rank on it.

To informationally promote their text in digital space research writers shall be properly instructed, provided the demand for keyword list is actually meant as that. Unfortunately, they receive, as mentioned above, no other regulation except the limit set for their number. The Internet advice for choosing keyword candidates is site-centric and caters for the interests of a very broad target audience[3]. Objective editorial recommendations are found yet in [Abramov 2011], who suggests that candidates for a keyword list shall:

- be chosen from terms found in the title, abstract, the opening and closing paragraphs of the text body;
- account for search engine statistics;

[3]See, for instancehttp://seomans.ru/nashel-info-about-plotnost-keywords.html (addressed 02.11.19)

Table 4: Keyword popularity according to search engine statistics

	Most Frequent Keyword *less frequent keyword **possible candidate for keyword	Keyword Density	Yandex Statistics (queries per month) date 19.11.19
Text 3	китайский язык (Chinese language) *иероглифика (hieroglyphics) **иероглифы (hieroglyphs)	0,46% 0,03% 0,52%	209586 2077 273552

- include terms and term collocations (e.g. *бухгалтерский учет основных средств, бухгалтерский учет, основные средства / fixed assets accounting, accounting, fixed assets*), cf.: "Longer search queries are narrower search queries, and narrower search queries are less competitive"[4];

- be not limited to 3-5 units, but include 10-15 ones.

These recommendations are definitely taking the job of promoting a research text as an informational product to be the author' personal issue. Submitting their text to a journal, authors have to "do as Romans do".

4 Discussion

The study of authors' keywords in 8 sample Russian research articles is based on our strong belief that compiling a keyword list is to a great degree unconventional writing practice within a highly conventional text format. The findings confirmed that the authors' choice of keyword candidates and their functioning in the text is rather arbitrary.

The quantitative description of keywords, which was meant to answer the question of how keywords show in the text and in what proportion, demonstrated that:

- keywords show unevenly in the text and demonstrate various degree of their presence, as well as a quite unspecific distribution among the text parts (synonymous texts); in some texts keywords are missing in the abstract, or both in the title and the abstract;

- the idea of "keyness" is maintained through the text by derivation (stemming), word composition / collocation and generalization / specification.

In the absence of editorial prescriptions keyword lists reveal a number of patterns underlying the actual composition of the keywords in the list, two of which seem rather significant:

[4]https://www.wordstream.com/blog/ws/2019/02/07/google-search-statistics (addressed 02.11.19)

- confirming the status of a minor text within the title – abstract – *keyword list - text body paradigm*, keyword sets show specific linguistic characteristics, namely integrity and cohesion, realized in their syntagmatic character (non-random, linear, a sort of "chain" structure) and thematic progression; this pattern is preferable in our sample texts;

- a keyword set may be composed as a logical sequence of classification / subject indexes of a definite subject area.

Keyword lists in the sample texts are rather a "linguistic" than informational image of a text (cf. a document search image), representing its content either according to subject classification headings, or suggesting a linear (chain) expansion of terms or article key issues.

The keywords informational value may be justified by the keyword ratio (keyword density) in the text, of which no author of the selected texts is presumably aware. Useful recommendations for promoting a research text as an informational product suggest picking up keyword candidates from terms used in the title, abstract, opening and closing paragraphs (so called "strong text positions"), accounting for search engine statistics, including terms and term collocations, increasing keywords number.

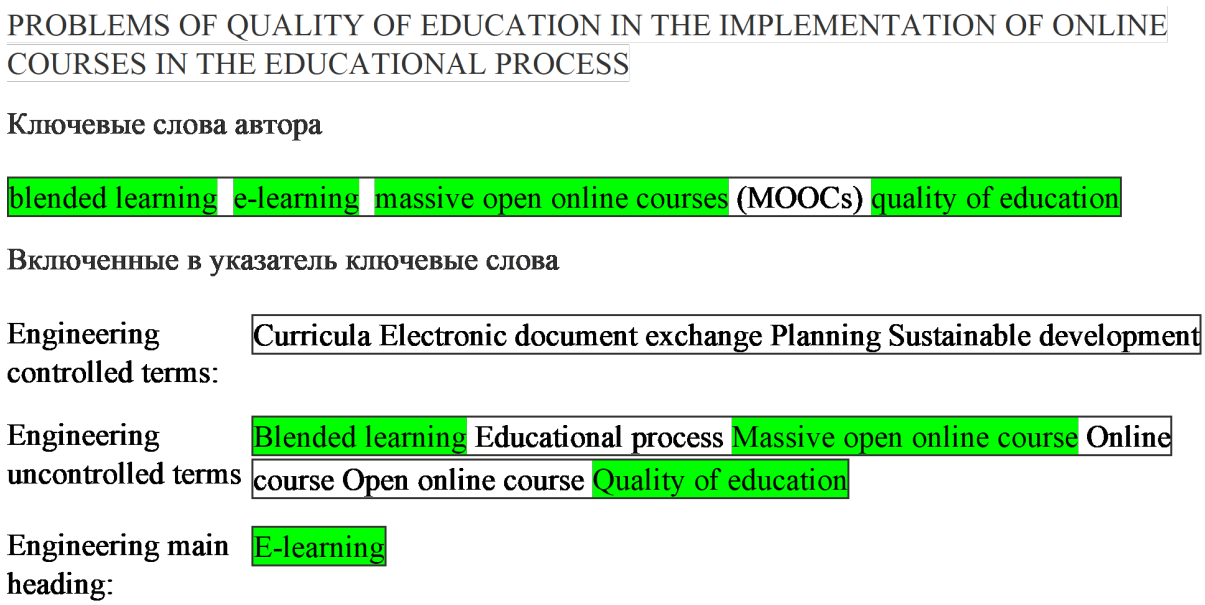
5 Conclusions

None of the 8 texts under study have a keyword list that can be treated as a text search image. To promote the text informationally in digital space the author shall be accordingly instructed, provided the demand for keyword list is actually meant as that. The last pointed research question in Introduction on whether there is any actual need for a keyword list is far from being an idle one.

In recent years authoritative publishers of research journals and peer-reviewed literature databases like Elsevier and Scopus ensure the policy of subject indexing which is performed for every text manually by a group of specially trained professionals. Subject indexing (see 3.2) is a process of indexing a text by human experts with keywords derived from the accepted system of controlled (authorized) terms (controlled vocabulary). Controlled vocabularies provide a way to organize knowledge for subsequent retrieval. "They are used in subject indexing schemes, subject headings, thesauri, taxonomies and other knowledge organization systems. Controlled vocabulary schemes mandate the use of predefined, authorized terms that have been preselected by the designers of the schemes, in contrast to natural language vocabularies, which have no such restriction". Elsevier admits, that authorized terms were manually added to 80% of Scopus publications. The result of subject indexing in Scopus is seen in figures 1 and 2. Highlighted are the matches of the authors' and added keywords.

Compared to automatic indexing, the use of a controlled vocabulary and human expert work can dramatically increase information retrieval progress Concerning the research question, one may notice that added controlled terms in both cases (figures 1 and 2) find no match in the authors' keyword lists, the latter matching the added uncontrolled terms. Whether it demonstrates the authors' incompetence or the need to update the controlled vocabulary in question, it is difficult to determine at present. One more thing worth noticing is that the choice of main heading in figure 1 seems correct, while the main heading defined in figure 2 seems erroneous. Anyhow, with subject indexing realized by human experts in such a way that it controls key terminology for different subject areas and promises dramatic increase in information retrieval, is the demand for author's keyword list still actual? Probably it isn't,

since there is a tendency not to include keywords into conference paper templates.


PROBLEMS OF QUALITY OF EDUCATION IN THE IMPLEMENTATION OF ONLINE COURSES IN THE EDUCATIONAL PROCESS

Ключевые слова автора

blended learning e-learning massive open online courses (MOOCs) quality of education

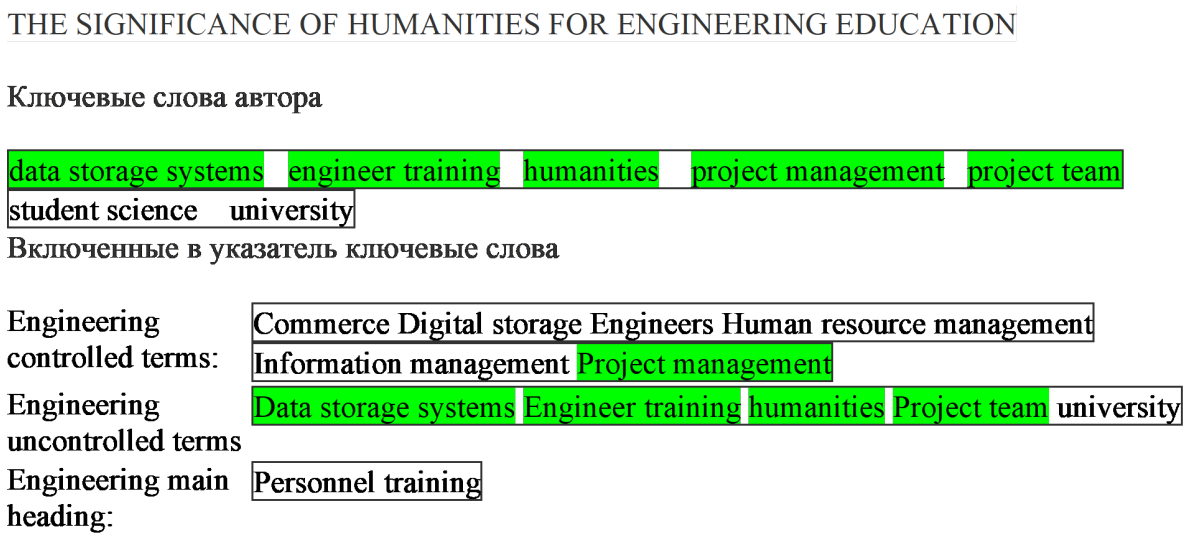
Включенные в указатель ключевые слова

Engineering controlled terms: Curricula Electronic document exchange Planning Sustainable development

Engineering uncontrolled terms: Blended learning Educational process Massive open online course Online course Open online course Quality of education

Engineering main heading: E-learning

Figure 1: Controlled terms added to author's keywords in "The Significance of Humanities for Engineering Education"


THE SIGNIFICANCE OF HUMANITIES FOR ENGINEERING EDUCATION

Ключевые слова автора

data storage systems engineer training humanities project management project team student science university

Включенные в указатель ключевые слова

Engineering controlled terms: Commerce Digital storage Engineers Human resource management Information management Project management

Engineering uncontrolled terms: Data storage systems Engineer training humanities Project team university

Engineering main heading: Personnel training

Figure 2: Controlled terms added to author's keywords in "Problems of Quality of Education in the Implementation of Online Courses in the Educational Process"

References

[Abramov 2011] Abramov, E.G. (2011) Selection of Keywords for a Research Article // Research Periodicals: Problems and Decisions. № 2. Pp. 35-40. (In Rus.) = Abramov E.G. Podbor klyuchevy‘kh slov dlya nauchnoj stat‘i // Nauchnaya periodika: problemy‘ i resheniya. 2011. № 2. S. 35-40.

[Arnold 1999] Arnold, I.V. (1999) Semantics. Stylistics. Intertextuality / Bukharkin, P.E. – Ed. SPb. 1999. – 444 p. (In Rus.) Arnol‘d I.V. Semantika. Stilistika. Intertekstual‘nost. : sb. St. / nauch. Red. Bukharkin P. E. SPb. : Izd-vo S. Peterb. Un-ta, 1999. – 444 s.

[Baeza-Yates, Ribeiro-Neto 2011] Baeza-Yates R., Ribeiro-Neto B. (2011) Modern Information Retrieval. The Concepts and Technology behind Search. Second edi¬tion. Pearson Education Ltd., Harlow, England, 2011. - 766 s.

[Bates, Maack 2010] Bates, M.J. and Maack, M.N. (eds.). (2010). Encyclopedia of Library and Information Sciences. Vol. 1-7. CRC Press, Boca Raton, USA.

[Bolotnova 2009] Bolotnova, N.S. (2009) Philological Analysis of Text. 4th ed. M. Flinta : Nauka. – 520 p. (In Rus.) = Bolotnova N.S. Filologicheskij analiz teksta 4 e izd. — M. : Flinta : Nauka, 2009. – 520 s [Firth 1957] Firth J.R. (1957) The Technique of Semantics // Papers in Linguistics 1934 – 1957. London. Oxford University Press. Pp. 7-33

[Busa 1980] Busa, R. (1980). The Annals of Humanities Computing: The Index Thomisticus. In Computers and the Humanities. 14(2). Pp. 83-90.

[Iagounova 2008] agounova, E.V. (2008) Set of Recognizable Words as Compression Texts (With Comparison of Key-Word Set) // Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue”. Issue 7 (14). Pp. 588-594. (In Rus.) = Yagunova, E.V. Nabor oporny‘kh slov kak vid svertki teksta (v sopostavlenii s naborom klyuchevy‘kh slov) // Komp‘yuternaya lingvistika i intellektual‘ny‘e tekhnologii. Materialy‘ mezhdunarodnoj konferenczii «Dialog - 2008». Vy‘pusk 7 (14). 2008. Pp. 588-594.

[Lukin 1999] ukin, V. A.(1999) Linguistic Theory basics for Text Analysis. M. Os-89. – 192 p. (In Rus.) = Lukin, V. A. Khudozhestvenny‘j tekst. Osnovy‘ lingvisticheskoj teorii. Analiticheskij minimum. M. Os‘-89. 1999. – 192 s.

[Lindemann, Kliche, Heid 2018] Lindemann D., Kliche F., Heid U. (2018). LexBib: A Corpus and Bibliography of Metalexicographical Publications. // Lexicography in Global Contexts. Proceedings of the XVIII EURALEX International Congress. – Ljubljana: University of Ljubljana, Centre for language resources and technologies, 2018. Pp.764–777

[Manning, Raghavan, Schutze 2008] Manning C. D., Raghavan P., Schutze H. (2008) Introduction to Information Retrieval Cambridge University Press New York, NY, USA. [Murzin, Stern 1991] Murzin L.N., Stern A.S. Text and Text Perception. Sverdlovsk. – 171 p. (In Rus.) = Murzin L.N., Shtern A.S. Tekst i ego vospriyatie. Sverdlovsk, Izd-vo Ural. un-ta, 1991 –171 s.

[Oxford Dictionary 2010] The Oxford Dictionary of English. 3rd edition. (2010). Ed. By A. Stevenson. Oxford University Press.

[Passonneau 2006] assonneau R. J. (2006). Measuring agreement on set-valued items (MASI) for semantic and pragmatic annotation. In Proceedings of the 5th International Conference on Language Resources and Evaluations (LREC 2006). Pp. 831–836.

[Ripp, Falke 2018] ipp, S., Falke S. (2018) Analyzing User Behavior with Matomo in the Online Information System Grammis. Lexicography in Global Contexts. Proceedings of the XVIII EURALEX International Congress. – Ljubljana: University of Ljubljana, Centre for language resources and technologies. Pp. 94–108

[Sakharny et al. 1984] Sakharny, L.V., Sibirsky, S.A., Stern, A.S. (1984) A Set of Key Words as a Text // Psycho-pedagogical and linguistic problems of text study. Perm. Pp. 81-83. (In Rus.) = Sakharny‘j L.V., Sibirskij S.A., Shtern A.S. Nabor klyuchevy‘kh slov kak tekst // Psikhologo-pedagogicheskie i lingvisticheskie problemy‘ issledovaniya teksta – Perm‘, 1984. S. 81-83.

[Sakharny, Stern 1988] Sakharny, L.V., Stern, A.S. (1988) A set of key words as a Text Type // Lexical aspects in professionally oriented teaching of foreign speech activity. Perm. Pp.34-51. (In Rus.) = Sakharny‘j L. V., Shtern A. S. Nabor klyuchevy‘kh slov kak tip teksta // Leksicheskie aspekty‘ v sisteme professional‘no-orientirovannogo obucheniya inoyazy‘chnoj rechevoj deyatel‘nosti. – Perm‘, 1988. S. 34-51.

[Shmelev 2002] Shmelev, A.D. A Russian Language World Model. Materials for a Dictionary. – 224 p. (in Rus.) – Russkaya yazy‘kovaya model‘ mira. Materialy‘ k slovaryu. Moskva: Yazy‘ki slavyanskoj kul‘tury. 2002. – 224 s.

[Stepanov 1997] Stepanov, Y.S. (1997) Constants. A Dictionary of Russian Culture. A Study. M. – 824 p. (In Rus.) = Konstanty‘. Slovar‘ russkoj kul‘tury‘. Opy‘t issledovaniya. M. Yazy‘ki russkoj kul‘tury. 1997. – 824 s.

[Stubbs 1996] Stubbs, M. (1996) Text and Corpus Analysis: Computer-assisted Studies of Language and Culture. Oxford, UK Cambridge, MA: Blackwell

[Stubbs 2001] Stubbs, M. (2001) Words and Phrases. Corpus Studies of Lexical Semantics. Oxford, UK: Blackwell

[Wierzbicka 1997] Wierzbicka, A. (1997) Understanding Cultures through Their Key Words: English, Russian, Polish, German, and Japanese (Oxford Studies in Anthropological Linguistics. No 8.

[Williams 1983] illiams R. (1983) Keywords: A Vocabulary of Culture and Society. New York. Oxford University Press. – 341 p.

[Zemskaya 1996] Zemskaya, T.A. (1996) Active Processes in Modern Wordbuilding // Russian at the End of XXth century. M. 1996. Pp. 90-141. (In Rus.) = Zemskaya T.A. Aktivny‘e proczessy‘ sovremennogo slovoproizvodstva // russkij yazy‘k koncza XX stoletiya. M. Yazy‘ki russkoj kul‘tury‘. 1996. S. 90–141

[Zeni et al. 2007] Zeni, N., Kiyavitskaya, N., Mich, L., Mylopoulos, J., Cordy, J. R. (2007). A Lightweight Approach to Semantic Annotation of Research Papers. // Natural Language Processing and Information Systems, Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. Pp. 61–72.