

Mathematical Model of Semantic Search and Search Optimization

Taras Basyuk ^[0000-0003-0813-0785], Andrii Vasyliuk ^[0000-0002-3666-7232],
Vasyl Lytvyn ^[0000-0002-9676-0180]

Information Systems and Networks Department,
Lviv Polytechnic National University, Lviv, Ukraine
Taras.M.Basyuk@lpnu.ua¹, Andrii.S.Vasyliuk@lpnu.ua²,
Vasyl.V.Lytvyn@lpnu.ua³

Abstract. This article analyzes the existing technologies of semantic search, which are used by search engines and outlines the main tasks that arise in this case. It is shown that for the description of search engine optimization algorithms, it is expedient to use a unified mathematical apparatus, in which algebra of algorithms is chosen. The result of the study is a synthesized model that allows to evaluate the content of the online resource for the purpose of similarity of texts and describes the process of forming ontology concepts to evaluate the possibilities of semantic information search. Also, it was formulated recommendations that must be followed in the process of search engine optimization using semantic search technology. The conducted research creates preconditions for designing the corresponding software units, their verification and adaptation to functioning in the global network.

Keywords: Internet resource; popularization; semantic search; search engines, algebra of algorithms.

1 Introduction

Search engines provide users with access to relevant information, highlighting it from a variety of online resources, the number of which continuously grows year by year. According to Netcraft, at the beginning of 2019, the number of websites was about 1.3 bln. The typical search engine's job is to find the query keywords and exclude the content analysis filling [1]. The situation is becoming even more complicated because of the fact that in each language there are such concepts as synonymy (words with identical meanings) or polysemy (words with several meanings), which greatly increases the number of irrelevant results. From this point of view, there arises the task of a detailed analysis of the content of Internet resources in order to minimize such situations.

Semantic web technologies provide a variety of tools for solving this problem by simplifying search using standardized ontological languages and using semantic search technologies that apply ontologies to store databases [2].

In view of this, the actual task of the study is to conduct research in the field of search engine optimization using semantic search technologies. At the same time, it is expedient to use a unified mathematical apparatus to display (describe) algorithms for the promotion of Internet resources. As such, it is proposed to use algebra of algorithms [3]. The aforementioned algebra of algorithms provides means for synthesis and minimization of algorithm algebra formula, which in the future allows to execute the synthesis of mathematical support on the basis of the algebra of algorithms operations properties and their transformations. According to the mentioned features, the use of algebra of algorithms is proposed as a means for creating a mathematical support for the process of search optimization using semantic search technologies.

1.1 Analysis of recent researches and publications

Nowadays, search engines use a relevant model to evaluate the accordance of the search query to the desired document, which in most cases can not cope with the tasks. This is primarily due to the approach used and the evaluation of the artificial criteria, such as the location of words on the page, their number, etc. [4,5].

The analysis of well-known research has shown [1,6] that the most popular technologies used in the process of finding relevant information are: boolean search - a combination of elements that allows to include and exclude from search results documents that contain certain words with the help of logical operators <and> (there are two or more elements or phrases), <not> (one item or phrase is excluded), <or> (one of the elements must be in the description) [6]; *Wildcard characters search* - involves the use of special characters ("*", "?"), which are used to replace the letters while writing [7]; *distance search* - displays documents that contain keywords which are at a certain distance and it is activated by using a tilde sign ("~") [8]; *"inaccurate search"* - provides finding pages that can match the search argument, even if the latter is inaccurately identical with the information sought, for example, the inaccurate comparison system may perform the correction of mistakes made while typing [9]; *contextual search* - defines the meaning of a word depending on the context of a page, rather than a single word (inaccurate search) and is the basis of Crystal Semantics Textonomy [10].

In view of the conducted analysis, it can be assumed that the set of used approaches can not claim to be versatile and search optimization remains an actual task. Especially in conditions of increasing spread of semantic search technologies which requires the application of new approaches and methods. From this perspective, the research task is to develop a mathematical support for the optimization of Internet resources for semantic search using the algebra of algorithms.

1.2 The main tasks of the study and their significance

The purpose of the research is to create a mathematical support for the process of search engine optimization of Internet resources using semantic search technologies. The conducted study will provide a means for promoting both existing and new online

resources, give an increase in the number of targeted visitors, and hence increase in conversions. To achieve this goal, the following main tasks need to be solved: analyze the existing semantic search technologies used by search engines and identify the main tasks that arise in this process, synthesize the models that can be used to evaluate the content of the online resource for the purpose of similarity of texts, formulate recommendations which need to be followed in the process of search optimization using semantic search technology.

The results of the research solve the actual problem of creating a mathematical support for the process of search engine optimization of Internet resources using semantic search technologies.

2 Major research results

The term semantic search is used to describe the attempts made by search engines to understand user queries. However, it is much wider than the normal search and includes the context in which the user is at the moment the search query is entered. For example, if the latter enters the word "university", and the previous request was "Lviv", then there is a probability that he is looking for information about universities in Lviv. In addition, the essential condition for contemporary semantic search is the use of the concept of entities, which is to associate with people, events or places [11]. In other words, the Lviv Polytechnic National University is an object characterized by the address, the number of buildings, the variety of institutes, directions and specialties, etc. Therefore, in the case of a search on the request of Lviv Polytechnic National University, and then - "which current specialties", the search engine displays the specialty of this university in the search results.

The need to use the concept of essences primarily involves the use of voice search technology in which an important feature is not only the "understanding" of the user's request, but also the definition of its "intention" in order to issue the most relevant result. In view of this, the approaches to optimizing Internet resources are radically changing. Namely, instead of writing content using competitive keywords, you first need to analyze the user's desire and create content that is relevant in nature and will answer the questions of the target audience [12]. Using this approach will facilitate the search engine promotion of the projected Internet resource not only for classical text queries, but also with the use of semantic search technologies.

Semantic search emerged from the notion of a semantic network, which is mainly based on ontologies, which in the context of computer science defines a set of entities (classes, attributes, and relationships), through which the domain modeling is carried out. Since they do not depend on lower-level data models, ontologies are used in the process of integrating heterogeneous databases, which provides tools for analyzing specific queries based on the relationships of related factors [13].

A conducted analysis of Google's search engine showed that the content-handling methods used in spinning texts are relatively easily recognized by the search engine with the use of mechanisms: Latent Semantic Indexing (a method of indexing websites in which searches take into account the overall content of the text, and not its

saturation with the keywords), latent Dirichlet allocation, which allows estimating the probability of the appearance of documents or terms beyond the text collection and for identifying the parameters in which the referencing and the term "frequency-inverse document frequency" is a term in which, for the purpose of determining the similarity between the texts for each pair, "the word of the current text is the text with which the comparison is made," the frequency of the occurrence of the word in the given text is found at the same time with the finding of the reverse frequency document). This indicates that it is impossible to apply classical approaches to search optimization by identifying the search statistical features of the repetition of words in a particular context and creating semantic correlations that are used in pertinent relevance technologies [14].

The conducted analysis of the mechanisms used by search engines to find relevant responses enables us to formulate an algorithm according to which it is possible to evaluate the content of an online resource for the purpose of comparing texts with other resources. In the future, the results obtained will be used in the process of constructing ontology concepts, which will provide tools for evaluating the content of the online resource in accordance with the requirements of semantic search.

At the first stage, pre-processing of the text is carried out, namely its transformation into the form of the data vector. Further elaboration consists in carrying out a training operation (cutting off endings and suffixes of words) and excluding non-informative phrases. The use of sedation methods is a widespread phenomenon in the global network and is widely used by search engines to evaluate the similarity of texts and the issuance of relevant information [8]. As of today, the literature describes a number of stemmers that perform morphological analysis (Stemka, MyStem, Pymorphy) or do word clipping (Porter stemmer, Paice / Husk Stemmer), but in most cases they are localized in certain languages to which the Ukrainian does not belong. In view of this, it is proposed to use the an improved method described by Golub T. [15] as a stemmer in this approach based on the modification of the Porter algorithm [16] and does not require the use of generated databases that reduces requirements for both hardware and to the number of calculations performed.

The synthesis of the modified Porter Stemmer algorithm formula was carried out in three stages: synthesis of sequences, synthesis of eliminations and minimization of the algorithm [17,18].

Synthesis of sequences. It is necessary to describe the following uniterms: R - reading the word, N - translating the character into lowercase, $D(a)$ - removing the apostrophe, $D(s)$ - removing the part of the word from the vowel-consonant, $D(z)$ - deleting the ending, $D(g)$ - deleting the vowel, $D(p)$ - removing one consonant, $D(m)$ - deleting the soft character. The considered algorithm contains 30 sequences. Each of them describes the following processes: $S1$ - execution of the algorithm in the case of no apostrophe, ending, loud at the end of the word, duplicate vowels and soft sign, $S2$ - the same cases as in $S1$, except that the word is an apostrophe, $S3$ - in the word there is an end, $S4$ - in the word there is an apostrophe and an end, $S5$ - under all conditions described in $S1$, in the word there is only a vowel at the end of the word, $S6$ - similar to $S5$, but in the word there is an apostrophe, $S7$ - in the word there is ending and loud at the end of the word, $S8$ - in the word there is a vowel at the end, the ending and

apostrophe, *S9* - in the word there is only a double consonant, *S10* - in the word there is a double consonant and apostrophe, *S11* - in the word there is double the consonant and the ending, *S12* - in the word there is a double consonant, the ending and the apostrophe, *S13* - in the word there is an end, at the end it is loud and doubly consonant, *S14* - in the word is the ending, at the end of the loud, doubly consonant and apostrophe, *S15* - in the word is at the end of the loud and doubly consonant, *S16* - similar to *S15*, except that the word is an apostrophe, *S17* - in the word there is a soft sign in the end, *S18* - in the word is a soft sign in the end and the apostrophe, *S19* - in the word there is an end and a soft sign, *S20* - in the word there are endings, *S21* - in the word there is a soft sign in the end, vowel in the end, *S22* - in the word there is a soft sign at the end, vowel at the end and an apostrophe, *S23* - in the word there is a double consonant and m ' which sign is at last *S24* - in the word is double consonant, soft sign at the end and apostrophe, *S25* - in the word there is an ending, loud at the end, doubly consonant and soft sign in the end, *S26* - in the word there are all cases, *S27* - in the word is loud at the end, doubly consonant and soft sign in the end, *S28* - in the word there is a loud end, doubly consonant, soft sign at the end and the apostrophe, *S29* - in the word there is an ending, doubly consonant and soft sign in the end, *S30* - in the word there is an ending, doubly consonant, a soft sign at the end and an apostrophe. Below are the following sequences.

S1= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , W$ *S2*= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , W$ *S3*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , W$
S4= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , W$ *S5*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(g)}^{\text{}} , W$
S6= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(g)}^{\text{}} , W$ *S7*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , \overbrace{D(g)}^{\text{}} , W$
S8= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , \overbrace{D(g)}^{\text{}} , W$ *S9*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(p)}^{\text{}} , W$
S10= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(p)}^{\text{}} , W$ *S11*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , \overbrace{D(p)}^{\text{}} , W$
S12= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , \overbrace{D(p)}^{\text{}} , W$ *S13*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , \overbrace{D(g)}^{\text{}} , \overbrace{D(p)}^{\text{}} , W$
S14= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , \overbrace{D(g)}^{\text{}} , \overbrace{D(p)}^{\text{}} , W$ *S15*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(g)}^{\text{}} , \overbrace{D(p)}^{\text{}} , W$
S16= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(g)}^{\text{}} , \overbrace{D(p)}^{\text{}} , W$ *S17*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(m)}^{\text{}} , W$
S18= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(m)}^{\text{}} , W$ *S19*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , \overbrace{D(m)}^{\text{}} , W$
S20= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , \overbrace{D(m)}^{\text{}} , W$ *S21*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(g)}^{\text{}} , \overbrace{D(m)}^{\text{}} , W$
S22= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(g)}^{\text{}} , \overbrace{D(m)}^{\text{}} , W$ *S23*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(p)}^{\text{}} , \overbrace{D(m)}^{\text{}} , W$
S24= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(p)}^{\text{}} , \overbrace{D(m)}^{\text{}} , W$ *S25*= $\overbrace{R, N}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , \overbrace{D(g)}^{\text{}} , \overbrace{D(p)}^{\text{}} , \overbrace{D(m)}^{\text{}} , W$
S26= $\overbrace{R, N}^{\text{}} , \overbrace{D(a)}^{\text{}} , \overbrace{D(s)}^{\text{}} , \overbrace{D(z)}^{\text{}} , \overbrace{D(g)}^{\text{}} , \overbrace{D(p)}^{\text{}} , \overbrace{D(m)}^{\text{}} , W$

$$\begin{aligned}
 S27 &= \overbrace{R, N}^{\text{---}} , \overbrace{D(s), D(g), D(p), D(m)}^{\text{---}} , W \\
 S28 &= \overbrace{R, N}^{\text{---}} , \overbrace{D(a), D(s), D(g), D(p), D(m)}^{\text{---}} , W \\
 S29 &= \overbrace{R, N}^{\text{---}} , \overbrace{D(s), D(z), D(p), D(m)}^{\text{---}} , W \\
 S30 &= \overbrace{R, N}^{\text{---}} , \overbrace{D(a), D(s), D(z), D(p), D(m)}^{\text{---}} , W
 \end{aligned}$$

After completing the substitution of the sequences and minimizing the algorithm, we obtain the following formula of the modified Porter stemmer algorithm:

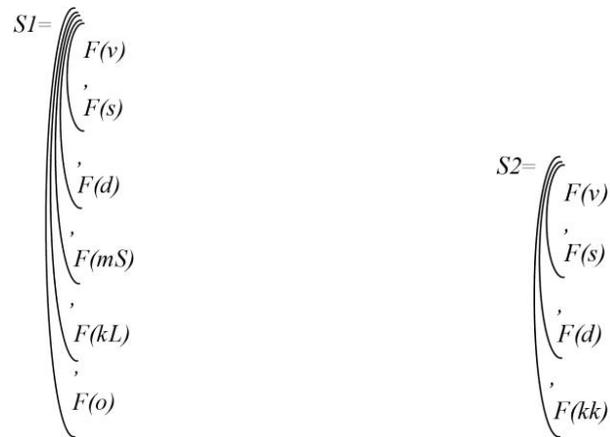
$$\begin{array}{l}
 \varnothing^i \\
 \hline
 \left(\begin{array}{l} R \\ , \\ N \end{array} \right) ; K ; u(i,n) - ? \\
 \left(\begin{array}{l} \overbrace{D(a) ; D(s) ; u_1?} \\ \overbrace{D(z) ; \overbrace{D(g) ; \overbrace{D(p) ; \overbrace{D(m) ; W ; u_5?} ; u_4?} ; u_3?} ; u_2?} \end{array} \right)
 \end{array}$$

The next stage involves removing the so-called stop words from the generated vector. Stop words are words that do not carry a content load but without them it is impossible to construct meaningful content. These include prepositions, pronouns, exclamations, punctuation marks, etc. [19]. As search engines are continuously improving, the word-of-mouth-recordings change as well, given the fact that a constant condition for updating and calculating their relation to the total number of content words. A significant number of stop words in the text negatively reflects on its evaluation by the user and creates the effect of meaningless content. The reverse situation, when the text includes not enough stop-words (creation of content oriented solely on search engines) also negatively affects readability and provokes the lack of interest in the user.

The next step is to determine the similarity of this text with the standard. To determine the degree of similarity between texts, it is proposed to use the statistical measure TF-IDF [10], which determines the frequency of occurrence of the word in this content. Next, the selection of the most meaningful words (key words) in the content is formed, which form the object, subject and predicate with the formation of possible patterns of searches / answers [20-23]. In this case, the words found will be displayed as an ordered list with links to the text paragraphs where they occur. In order to deduce the complete information at this step, the word lexeme is displayed, indicating the objects and indicating the concept of them. The lexical value reflects distinctive, individual features of the subject. It is proposed them to be output by displaying the original with automatic positioning on the text fragment found and keyword selection. The final stage of the algorithm work is the construction of an

ontology, in which the definition of classes and their hierarchy is carried out. Next, the properties of each class, the restrictions and types of properties values are determined. The result of this step is the set of concepts and relationships in the form of triplets that conforms to the RDF (Schema) standard and provides the ability to translate them into the OWL language [13]. For the convenience of evaluating the construction of the ontological model, it is proposed to implement using the OntoViz module. Upon completion of the construction, it is proposed to use the FaCT ++ consideration module to identify possible non-conformities in the ontology and to compare it (in the long run) with the linguistic base for the Ukrainian language (Ukrainian WordNet) [24], which will help to assess the completeness of content coverage and its relevance to the requirements of semantic search.

To synthesize the formula for a search engine optimization resource under a semantic search, one must describe the following uniterms: $F(v)$ - create a data vector, $F(s)$ - perform a training operation, $F(d)$ - remove stop words, $F(mS)$ - measure the size similarity, $F(kL)$ - forming a list of key and displaying tokens, $F(o)$ - constructing the model's ontology and visualization, and $F(kk)$ creating "useful content". Linear actions are described in sequences $S1$ and $S2$:



On the condition of checking whether the permissible value of the ratio of total content to the stop words, these sequences are eliminated by elimination of LI .

$$LI = \left| \frac{S_2}{S_1} ; u_1 ? \right|$$

After substituting the corresponding sequences in the elimination, we obtain the following formula of the algorithm:

$$LI = \left(\left(\left(F(v) \right) , \left(F(s) \right) \right) , \left(F(d) \right) \right) ; \left(\left(\left(F(v) \right) , \left(F(s) \right) \right) , \left(F(d) \right) \right) ; \left(F(kk) \right) ; \left(\left(\left(F(mS) \right) , \left(F(kL) \right) \right) , \left(F(o) \right) \right) ; u_i ?$$

As a result of minimizing the algorithm formula by the number of uniterms, we obtain the formula for search optimization of the resource under the semantic search.

$$LI = \left(\left(\left(F(v) \right) , \left(F(s) \right) \right) , \left(F(d) \right) \right) ; \left(F(kk) \right) ; \left(\left(\left(F(mS) \right) , \left(F(kL) \right) \right) , \left(F(o) \right) \right) ; u_i ?$$

As you can see, semantic search is extremely important in the process of conducting an SEO company. In view of this, the analysis of known strategies made it possible to formulate recommendations that should be followed in the process of search engine optimization using semantic search technologies:

Creating quality content. Modern search engines implement artificial intelligence methods in order to provide a possible dialogue with the user. To perform this function, they need a large array of information, landmarks, expert content. From this point of view, it is necessary to create authoritative content in the relevant subject area, to become a source of expert information so that search engines can refer to a popularized resource.

Orientation to the answer. A necessary condition is content creation focusing on questions / answers. The research conducted showed that search engines prefer to

display information in the form of numbered lists or step-by-step instructions that respond to users' questions and begin with the words "how to do", "why", "what," and so on.

Technical structuring of content. Structuring data for markup is to annotate the pages of the online resource, making them understandable to search engines. Using structured markup not only gives search engines the opportunity to better understand the content, but also improves search quality by displaying results in a snippet (zero position) that gives the user additional information about the content on the page and improves the Click-through rate using semantic search. It is advisable to verify the technical structure using the Structured Data Testing Tool.

Use of internal links. Internal links played and continue to play a significant role in creating a positive user experience by providing navigation as an online resource. In doing so, you need to link landing pages, add contextual links to important content elements, prevent the occurrence of broken links, etc.

3 Conclusion

As a result of the research, the existing semantic search technologies used by search engines are analyzed and the main problems that arise here are identified. Finished synthesis of models according to which it is possible to evaluate the content of the online resource for the similarity of texts with other resources and describes the process of forming ontology concepts to evaluate the possibilities of semantic information search. Unlike the classical tools, it provides the means to minimize them by the number of uniterms and study the corresponding mathematical models. The recommendations are to be followed in the process of search engine optimization using semantic search technologies.

Further research will focus on the design of relevant software units, their verification and adaptation to the operation of the global network.

References

1. Grappone, J.: Search Engine Optimization (SEO): An Hour a Day. In: United States, Wiley Publishing. (2013)
2. Su, J., Sachenko, A., Lytvyn, V., Vysotska, V., Dosyn, D.: Model of Touristic Information Resources Integration According to User Needs. In: International Scientific and Technical Conference on Computer Sciences and Information Technologies, 113-116 (2018)
3. Ovsyak, V.: Algorithms: methods of construction, optimization, probability research. In: Lviv, Svit. (2001) (In Ukrainian)
4. Basyuk, T.: Popularization of website and without anchor promotion. In: International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT), 193-195 (2016)
5. Basyuk, T.: Innerlinking website pages and weight of links. In: International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT), 12-15 (2017)

6. Amerland, D.: Google Semantic Search: Search Engine Optimization (SEO) Techniques That Get Your Company More Traffic, Increase Brand Impact, and Amplify Your Online Presence. In: United States, Que Publishing. (2013)
7. Vysotska, V., Basto Fandes, V., Lytvyn, V., Emmerich, M., Hrendus, M.: Method for Determining Linguometric Coefficient Dynamics of Ukrainian Text Content Authorship. In: International Conference on Computer Science and Information Technologies (CSIT), 132-151 (2019)
8. Najman, L., Talbot, H.: Mathematical Morphology: From Theory to Applications. In: United Kingdom, Wiley-ISTE. (2010)
9. Frank, Y.: Shih Image Processing and Mathematical Morphology: Fundamentals and Applications. In: United States, CRC Press. (2009)
10. Jones, K.: A statistical interpretation of term specificity and its application in retrieval. In: Journal of Documentation, vol. 60(5), 493-502 (2004)
11. Basyuk, T.: The Popularization Problem of Websites and Analysis of Competitors. In: Advances in Intelligent Systems and Computing II (CSIT), vol. 689, 54-65. (2018)
12. Bailin, A., Grafstein, A.: Readability: Text and Context. In: Palgrave Macmillan. (2016)
13. Gašević, D., Djuric, D., Devedžić, V., Selic, B., Bézivin, J.: Model Driven Engineering and Ontology Development. In: Springer. (2009)
14. Bast, H., Buchhold, B., Hausmann E.: Semantic Search on Text and Knowledge Bases (Foundations and Trends in Information Retrieval). In: US, Now Publishers Inc. (2016)
15. Golub, T., Tyagunova, Yu.: The method of Ukrainian language stemming for the classification of documents based on Porter's algorithm. In: Scientific works of the Donetsk National Technical University, vol. 1, 59-63 (2017) (In Ukrainian)
16. Porter, M.: An algorithm for suffix stripping Program. In: Data Technologies and Application, vol. 40(3), 211-218 (2006)
17. Vysotska, V., Fernandes, V.B., Emmerich, M.: Web content support method in electronic business systems. In: CEUR Workshop Proceedings, Vol-2136, 20-41 (2018)
18. Vysotska, V., Hasko, R., Kuchkovskiy, V.: Process analysis in electronic content commerce system. In: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT 2015, 120-123 (2015)
19. Vysotska, V., Kanishcheva, O., Hlavcheva, Y.: Authorship Identification of the Scientific Text in Ukrainian with Using the Lingvometry Methods. In: International Conference on Computer Science and Information Technologies (CSIT), 34-38 (2018)
20. Basyuk, T.: Popularization of Internet resources by using "featured snippets". In: International conference System Analysis and Information Technology, 190-191 (2018)
21. Korobchinsky, M., Vysotska, V., Chyrun, L., Chyrun, L.: Peculiarities of Content Forming and Analysis in Internet Newspaper Covering Music News, In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 52-57 (2017)
22. Kanishcheva, O., Vysotska, V., Chyrun, L., Gozhyj, A.: Method of Integration and Content Management of the Information Resources Network. In: Advances in Intelligent Systems and Computing, 689, Springer, 204-216 (2018)
23. Naum, O., Chyrun, L., Kanishcheva, O., Vysotska, V.: Intellectual System Design for Content Formation. In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 131-138 (2017)
24. Anisimov, A., Marchenko, O., Nikonenko, A., Porkhun, E., Taranukha, V.: Ukrainian WordNet: Creation and Filling. In: International Conference on Flexible Query Answering Systems (FQAS), 649-660. (2013)