DOI: https://doi.org/10.15276/aait.02.2019.1
UDC 004.912

METHOD OF SEARCHING TERM INTERPRETATIONS FOR DOMAIN DICTIONARIES, USED FOR DEVELOPING SOFTWARE

Oleksii B. Kungurtsev[1)]
ORCID: https://orcid.org/0000-0002-3207-7315, abkun@te.net.ua
Svetlana L. Zinovatnaya[1)]
ORCID: https://orcid.org/0000-0002-9190-6486, zinovatnaya.svetlana@opu.ua
Iana V. Potochniak[1)]
ORCID: https://orcid.org/ 0000-0003-1291-1146, yana.onpu@gmail.com
[1)] Odessa National Polytechnic University, 1, Shevchenko Avenue. Odesa, 65044, Ukraine

ABSTRACT

In this paper, a method of forming definitions of terms for a vocabulary of a subject domain using existing explanatory dictionaries is proposed. It is shown that with a combined search for terms and their interpretations, it is possible to find about ten percent of definitions, which is clearly not enough. A method of automated search for the interpretation of terms is proposed, involving the use of existing explanatory dictionaries. A mathematical model of the subject domain dictionary entry is proposed. A mathematical model of an explanatory dictionary entry is proposed, taking into account the headword, a variety of interpretations of the word, litters and stable phrases. A mechanism has been developed for extracting definitions of a term from an explanatory dictionary depending on the structure of its dictionary entry. An algorithm for automated search for definitions for single-word terms has been developed. An algorithm has been developed for the automated search for definitions for verbose terms, based on the selection of nouns from the term. A mechanism for assessing the quality of possible interpretations, depending on the occurrence of terms from the subject domain, is proposed. A mechanism has been developed for the choice of definitions, when the terms from the vocabulary of the subject domain and the explanatory dictionary are coinciding incompletely, which is based on the procedure of term decomposition, the search for partial interpretations and the synthesis of the resulting interpretation. The software developed that allows to organize the search for interpretations of terms both in local explanatory dictionaries (previously loaded into the system), and in online dictionaries. The expert's task includes the evaluation the interpretations found and possibly editorial correction of them. Experimental evaluation of the effectiveness of the use of a software product showed a reduction in the expert's working time compared to the "manual mode" by approximately four times.

Keywords: Domain Dictionary; One-Word Term; Multiword Term; Morphological Analysis; Mathematical Model of the Term; Interpretation of the Term; Text Document

***For citation*: Kungurtsev O. B., Zinovatnaya S. L., Potochniak Ia. V. Method of Searching Term Interpretations for Domain Dictionaries, Used for Developing Software. *Applied Aspects of Information Technology*. 2019; Vol. 2 No.1:11–19. DOI: https://doi.org/10.15276/aait.02.2019.1**

I. INTRODUCTION

Usually, a domain dictionary (DD) is a specialized explanatory dictionary, which gives definitions to a set of concepts related to the activities of a certain organizational structure [1, 2], [3]. Domain dictionary is used to solve a variety of tasks related to the creation and maintenance of software products (SP): formulation and coordination of requirements for SP, database design, creation of user interfaces, writing various manuals, etc. [4, 5], [6, 7].

There are two main types of dictionaries according to their content: encyclopedic and linguistic [5]. The object of the description in the encyclopedic dictionary and encyclopedia – various objects, phenomena and concepts;

the object of the description in the linguistic dictionary – a unit of language, most often a word.

By the way the material is organized, linguistic dictionaries are divided into alphabetic (most common), family (one dictionary entry interprets not the word, but the entire word-formation family) and semi-family (derived words of a different grammatical category than the headword).

The main type of the linguistic dictionary is the explanatory dictionary. Explanatory dictionaries differ in the volume of the dictionary, technical means of presenting the material. Therefore, before using the dictionary, it is necessary to be acquainted with the "System notes" – conditional abbreviations (usually found in the intro to the dictionary).

According to the functions and purposes of creating, explanatory dictionaries are divided into descriptive and normative.

Descriptive dictionaries are designated for a complete description of the vocabulary of a certain sphere and fixation of all available uses. ("Explanatory dictionary of the common great Russian language" V. I. Dal [8, 9], [10], "Explanatory dictionary of the English language" Oxford Dictionaries Online [11], "Explanatory dictionary of the Ukrainian language" Institute of Linguistics Online A. A. Potebni [12]).

The purpose of the normative dictionary is to show the standard use of the word, eliminating not only the wrong use of words associated with an erroneous understanding of their meanings, but also

those uses that do not correspond to the communicative situation.

The dictionary consists of dictionary entries [13, 14], [15, 16]. Dictionary entry is brief linguistic information about a word. In its turn, the dictionary entry consists of the following components:

– the headword (usually in bold and capital letters) with the emphasis. Sometimes it contains comments to the pronunciation in square brackets;

– interpretation of the word contains:

a) notes – a brief description the words expressed by the adopted reduction of the corresponding term describing the use of the word (usually in italics);

b) definition (dictionary definition);

– illustrative material as a means of word semantics;

– collocations based on the headword;

– derived words (a sign of the family dictionary).

Compilation of explanatory dictionary for a certain language or a certain branch of knowledge is a very time – consuming and poorly automated process that requires many months of work of highly qualified specialists. Therefore, when compiling the DD, there is a task to automate and speed up the process of dictionary compiling.

II. ANALYSIS OF THE LITERATURE DATA AND FORMULATION OF THE PROBLEM

There are a number of works on creation of DD in English [13], [11], Russian [17-18], [19], Ukrainian [20, 21], [22] and other languages.

The paper [1] focuses on the automated selection of terms from the texts in Russian from a given subject area. The interpretation of the terms is entrusted to the expert, who is offered to use a pre-selected set of dictionaries online. The study [7] deals with the automation of the allocation of terms from the texts in the Ukrainian language, but there are no solutions to automate the process of interpretation of terms.

A number of studies have considered in detail the structure of existing dictionaries and the conditions for the successful search of definitions, however the algorithms for their selection in the automated search is not proposed [5; 23]. In addition, the dictionary entry of the explanatory dictionary is mainly intended to define one word, while the terms of the subject area, as a rule, contain several words, for example, “operating system”, “waybill”, “medical history”, “schedule of commuter trains”.

There are certain rules for the interpretation of collocations based on the headword, but there are not enough of formalized rules for their allocation [24; 25]. Therefore, the automation of the retrieval of the interpretations of the terms is highly relevant and largely unresolved challenge [26].

The problem of interpretations highlighting is the high complexity of their definition [27].

The proposed solution is to automate the process of determining the definitions of terms for the DD based on the allocation of interpretations from the analyzed document and public dictionaries.

To solve the problem, the following tasks have been formulated:

– determination of the conditions for determining definitions directly from the analyzed document;

– allocation of definitions of the term from the dictionary;

– filtering of definitions depends on notes;

– filtering definitions based on the analysis of the entry of terms from the DD;

– selection and layout of definitions for multiword terms.

III. SEARCH OF INTERPRETATIONS COMBINED WITH A SELECTION OF TERMS FROM DOCUMENTS

The analyzed document [28] may introduce some new concepts (terms) or provide a new interpretation of the known ones. Then the definition of the term can be included in the text in close proximity to the term itself. To verify the effectiveness of the interpretation search directly in the text on the basis of which the DD is build, the analysis of text documents from various subject areas with a total volume of 50,000 words was carried out. As a result, 257 terms were allocated. For the 27 terms interpretation were found directly in the analyzed texts. On the basis of the study it was concluded that the development of methods of interpretation search combined with the selection of terms is futile.

IV. THE USE OF EXISTING EXPLANATORY DICTIONARIES

We assume that the search for definitions of terms in all cases will use pre-compiled specialized or broad-profile dictionaries.

At Fig.1 the dictionary, entry is presented in a simplified form.

If you want to find an interpretation of a one-word term, you should choose one or more definitions from *Definition1 ... Definition n* (for example, for the

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

term “System”). If the task is to find a definition for a multiword term (for example, “Operating system”), then you should look for a suitable collocation (in Fig. 1 is the system + T1, system+T2) and make a choice from the definitions that apply to it.

For the formation of DD, we will use a work piece for open DD without interpretations (contains only terms and their frequency characteristics), as well as some downloaded DD with online access, organized through some system that allows to extract information from sites such as Interpretation [29].

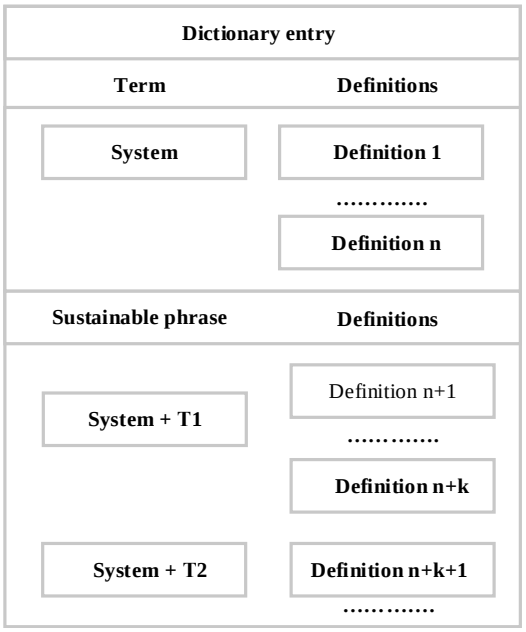

Fig. 1. **Simplified structure of the dictionary entry**
Source: **compiled by the authors**

Let us present the explanatory DD in the form of

$$To = \{< to, tt >\}, \quad (1)$$

where: *to* is a term from the subject area; *tt* – interpretation of the term (one or more sentences, originally an empty line– *tt*=«»).

Let’s present the explanatory dictionary of a wide profile in the form of:

$$Td = \{< td, ma >\}, \quad (2)$$

where: *td* is a term from the *Td* dictionary; *ma* – multiple interpretations of the term *td* (dictionary entry).

In general, the term *to* can consist of several words:

$$to = e_1 ... e_n. \quad (3)$$

Search for interpretations of one-word terms (OT).

We will assume beforehand that the term *to* is OT (*to* = e_1). Then the task of interpreting *to* can be formulated as follows. If *td* is found, such that *td* = *to*, then from a variety of interpretations *ma*, we need to choose one or more interpretation definition options (IDO), which for some formal features is most suitable for the subject area.

Let's set the task of reducing possible IDO. Typically, dictionaries contain abbreviations that define the scope of the term, for example, area of activity (Mat. – mathematics, mus. – music, comp. sc. – computer science), style (folk-poet – folklore-poetic, dismiss. – dismissive), etc. We call such reductions characteristics of the option. Then IDO interpretation of the term (family) can be represented as:

$$ma_i = \{< mc, tx >\}, \quad (4)$$

where: *mc* is the set of IDO characteristics; *tx* – IDO text.

Since not every IDO can contain characteristics, it is possible that *mc* = Ø.

For a particular domain, we will create many invalid *mCa* characteristics and lots of options for *mCi* characteristics. This allows us to distinguish from the set of *ma* a subset of ma' interpretations of the term that satisfy the following conditions:

$$\begin{matrix} \exists c_j \mid c_j \in mc \wedge c_j \in mCi \\ \forall c_j \mid c_j \in mc \wedge c_j \notin mCa \end{matrix}\}. \quad (5)$$

Since not every family has a characteristic and several different families can have the same characteristics, after selecting the IDO according to the conditions (4), there may be several of them. To further reduce the number of IDO, it is proposed to select the most suitable IDO from *ma'* by counting the number of occurrences of terms from *To* in each *tx* from *ma'*. To do this, let's introduce *ma'* as:

$$ma' = \{< tx, k >\},$$

where: *k* is the number of term occurrences from *To* to *tx*; *tx* is represented as a sequence of words (*tx* = $e_1 \ldots e_n$).

We write operation of calculation of the number of occurrences of term from *To* in *tx* (preliminary *k* = 0):

$$\begin{matrix} (if((e_i = to_j)e_i \in tx \wedge to_j \in To)k := k+1) \\ i = 1, n; j = 1, q \end{matrix}, \quad (6)$$

where: $q = |To|$.

After calculation *k* all IDO, which are not part of the terms of the subject area:

$$\forall tx_j \mid tx_j \in ma' \wedge k_j = 0,$$

excluded from ma', and the remaining are sorted in ascending order by *k*.

$$ma' = \{< tx_j, k_j >\} l = 1, m-1; k_l \geq k_{l+1},$$

where: $m = |ma'|$.

Thus, the expert is provided with *m* IDO to select and edit the term.

The search method of the multiword term

interpretations (MT).

There are dictionaries of word combinations [5], but they represent only a small number of commonly used word combinations. In ordinary explanatory dictionaries word combination are used as an IDO of some basic term. This determines the next stages of interpretation search for MT.

A) The allocation of the MT nouns. Since nouns carry the main semantic load, the interpretation of MT suggests the search for interpretation to start with nouns.

We introduce additional information into the MT submission in accordance with (3). To do this, each element (word) is represented as:

$$e_i =< w_i, w'_i, a_i >, \quad (7)$$

where: w_i represents one of the words of MT, w'_i – represents w_i in normalized form, a_i i - part of speech w_i (noun, adjective, number, ...).

B) Definition of search priorities for nouns. Some nouns that are included in MT can represent OT in the subject area under study. It is recommended to start the search for interpretations with these nouns. This will allow you to partially combine the search of MT and OT. In addition, it is possible to assume that a noun used only as a part of MT has a highly specialized meaning, which will make it difficult to find it in *Td*. Thus, we form a sequence of words to search for IDO in *Td.* :

$$to' = w'_1,...,w'_i,...,w'_n, \quad (8)$$

where: their location is subject to the following condition:

$$(\forall w'_i \mid (w'_i \in To)\exists w'_{i-1} \in To) \wedge (\forall w'_i \mid (a_i = noun)$$

$$\exists w'_{i-1} \mid a_{i-1} = noun); i = 2, n.$$

C) Search of OT in the *Td*. For each element of (8) satisfying the condition $w'_i \in To$, is searched according to IDO in paragraph 1.

D) Search of IDO for the MT based on noun. Dictionary entry *Td* can have a number of interpretations of collocations, which include the defined term. The generally accepted word order in such a phrase can be changed. Often defined term (represented as the first letter of the word) takes the first position in the word combination. In accordance with the above, it is proposed to present the explanatory dictionary entry of the *Td* in the form of:

$$ar =< td, \{s, ts\} >,$$

where: *S* is a word combination; *ts* is IDO, corresponding to the interpretation of the term *td* in this word combination.

Each phrase will present many of its constituent words, with previous deletion of linking words:

$$s = \{ws_i\} i = 1, n.$$

Similarly, let's present MT:

$$to = \{wt_j\} j = 1, k,$$

We define a set of *ma'* IDO for MT that can be provided to an expert. Originally *ma'* = Ø. Since there are, no assurance that in *Td* will be found the word combination matching with *to*, it makes sense to memorize the word combination, which partially correspond with *to*.

The terms of placement of *S* in *ma'* is:

$$(X = to \cap s) \wedge \mid X \mid \geq 2 \ , \ to \subseteq s \ .$$

Therefore, if an *S* containing two or more words from to is found, then such an incomplete interpretation is included in *ma'*. Also *ma '* includes all *S* that match with *to* or have additional words.

E) Assessment of the relevance of IDO from *ma'*. If one or more IDO are found, such that *to* = *S* then all $s \neq to$ are excluded from *ma'*.

If $|ma'| \geq 2$, then it is necessary to estimate each IDO entering *ma'*. To do this, in accordance with the operation (6) the occurrences of terms from a *To* in each IDO of the *ma'* must be calculated. Elements of the *ma'* set are sorted in descending order by the number of term occurrences. IDO that do not contain terms or have a small number of term occurrences are discarded.

F) Search for IDO based on all the words in the MT. If the search for IDO on the basis of nouns did not give results, in accordance with the previously described procedure, the search for IDO is based on all the words that make up the MT.

G) Search for IDO, which partially cover MT. If, as a result of the search for IDO in accordance with paragraphs D) and E) in the set of *ma'* there are no elements such that *to* = *S*, it is proposed to use phrases containing only part of the words included in MT.

We assume that the number of words in each element *ma'* is a measure of its correspondence to MT. To do this, we order the elements of *ma'* in descending order of their power:

$$\forall(s_i \in ma')\exists s_{i-1} \in ma' \mid (\mid s_{i-1} \mid \geq \mid s_i \mid); i = 2, n.$$

Similarly, to the point E) let's assess the relevance of IDO from *ma'*. IDO with a small number of words and a small number of occurrences of terms from *To* are discarded.

For the remaining IDO of *ma'* it is necessary to give the interpretation of words that are included in MT, but not included in *S* and those that are included in *S*, but not included in MT. We define sets of these words:

$$Y1 = to \setminus s \ \text{ where } \mid to \mid > \mid s \mid,$$

$$Y2 = s \setminus to \ \text{ where } \mid s \mid > \mid to \mid.$$

It is necessary to present one IDO as several constituent parts required for the introduction of structuring for the *S*. Let's present (IDO) as:

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

$$s = <\{ws_i\},\{w_l,\{w_j\}\}>, \quad (9)$$

where: $\{ws_i\}$ is the set of words representing the main partial interpretation of MT; w_l is one of the words in the set *Y1* or *Y2*; w_j is set of words representing the interpretation of the word w_l.

The representation of IDO in the form (9) allows providing the expert with the most comprehensive information available on options of the interpretation of MT.

V. METHOD IMPLEMENTATION

In accordance with the proposed method of determining the interpretation, algorithms to find the definition of one-word and multi-word terms were developed.

The search of a single word term interpretation is shown with following algorithm:

1. Acquiring information about the term (normal form of representing a term).

2. Performing a process of search an interpretation of the term in a local dictionary. If the interpretation was found, performing fixation of the interpretation, otherwise, performing search of the interpretation in connected dictionary.

3. Performing a return of the allocated term interpretations:

– if there was found a single interpretation of the term - performing save of the interpretation;

– if there was found several interpretations of the term – performing process filtering and sorting definitions based on data of the domain dictionary and then performing save of interpretations;

– if there was not found an interpretation of the term, returns only it name, which allows to an expert perform a search and save the interpretation manually.

The search of a multi-word term interpretation is shown with following algorithm.

1. Acquiring information about the term (the term, component parts of the term and normal form of all it parts).

2. Performing process of allocating keywords and combinations of partial phrases.

3. For every selected keyword performed the process of search a single word term and fixation of returned value.

4. Performing search of an interpretation for the single word term in either connected or local dictionary.

5. Performing fixation of the interpretation for the returned value.

6. For every fixed term and it interpretation, performing acknowledge of similarity with a multi word term by a set of words.

7. Performing a process of sorting the interpretations by the result of the a know ledges and then goes validation on whether the interpretation exceeds the limit of number returned definitions:

– if the limit is exceeded – performing exclusion of the interpretation by the least number of occurrence of the term in the domain dictionary.

– if the limit is not exceeded – exclusion is not performed.

8. Performing a return of the allocated term interpretations:

– if there was found a single interpretation of the term - performing save of the interpretation.

– if there was found several interpretations of the term – performing process filtering and sorting definitions based on data of the domain dictionary and then performing save of interpretations.

– if there was not found an interpretation of the term, returns only it name, which allows to an expert perform a search and save the interpretation manually.

On Fig. 2 is given schema of packages of program modules. Purpose of packages is following:

– User side – the user side, the user is “Expert”, which edits the interpretation of terms and stores them based on the results of the search.

– Terms allocation system – a term allocation system that provides a list of terms with parameters for searching.

– WebSites / SeleniumWebDriver – websites with online dictionaries that work with SeleniumWebDriver.

– WebDictionary is responsible for working with online dictionaries and looking for a term in them.

– Dictionary – is responsible for looking for the interpretation of the term, the main class of the program. It focuses on the basic logic of finding a term's interpretation.

– Term - the term model, describes the term and its parameters.

– FileWorker - is responsible for working with files (reading connected dictionaries, reading and writing the internal dictionary).

– View - view, responsible for the graphical component and work with the user interface.

On Fig. 3 given the program product which implements the algorithm of search of interpretations single word and multi word terms.

At Fig. 3 shown a window of the software product that represents the results of the search for the interpretation of the multiword term “computer software”. The word combination was not found, the result was presented in the form of interpretations for two words – “computer” and “software”.

At Fig. 4 the result of the search for the interpretation of the term “computer programs” is

presented. Of the two definitions of “computer” and “computer program”, preference was given to the second one based on the calculation of entries in the definition of other terms from the DD.

VI. EVALUATION OF THE EFFECTIVENESS OF THE METHOD AND SOFTWARE, DEFINITION OF TERMS OF INTERFERENCE

For the experiment with 63400 terms were randomly selected 5 lists of 30 terms from different subject areas. Tests were carried out in 3 modes: in manual mode, when the expert had to find and edit the definition of the term using online dictionaries; in the automated mode, when the search was performed by the program on the built-in and external dictionaries, and the expert edited the results; in the optimized mode, when the search was performed by the program, but the previously obtained terms from the corresponding subject area were taken into account. The results of the experiment are shown in Fig. 5

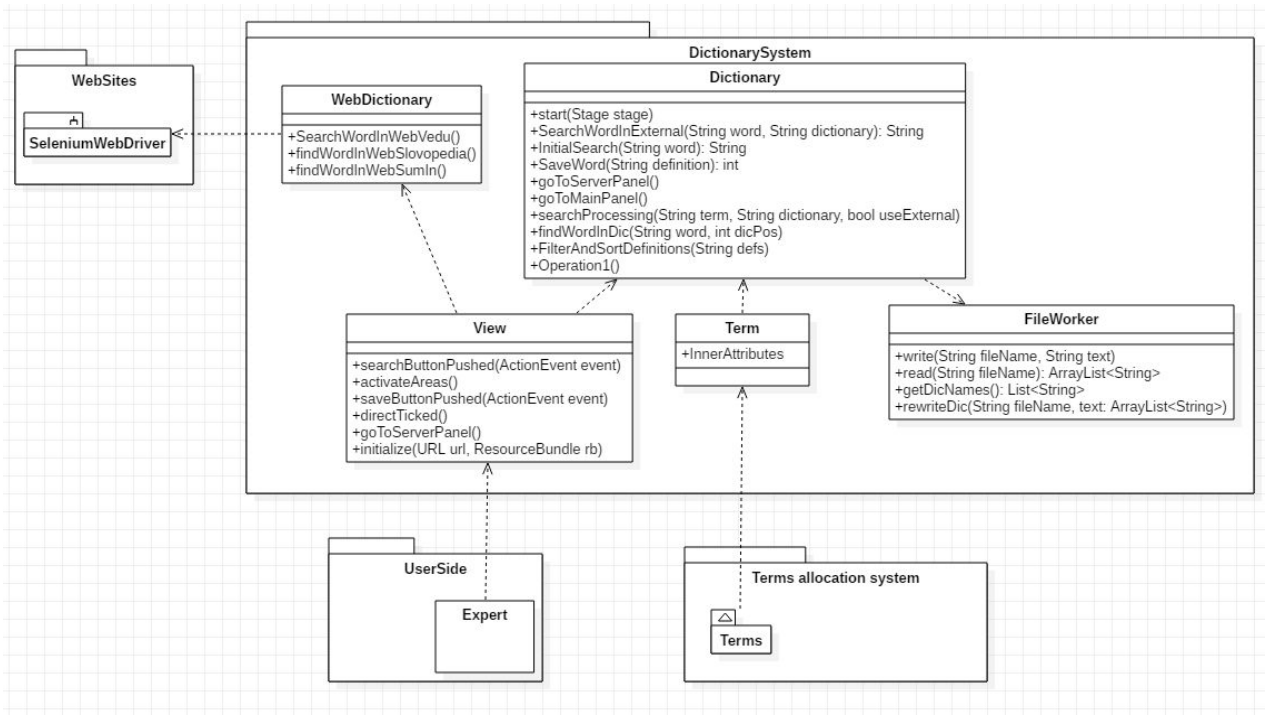

Fig. 2. **Schema of packages of program modules**
Source: **compiled by the authors**

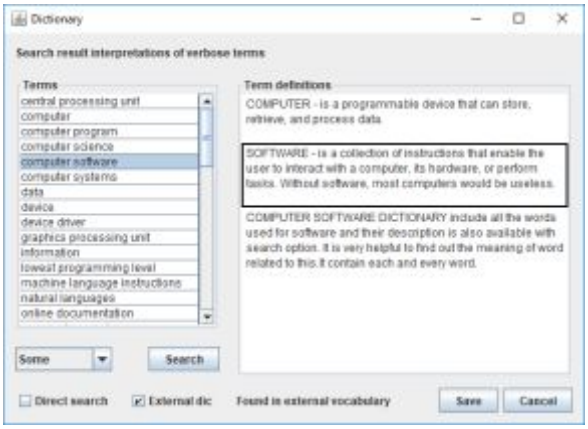

Fig. 3. **Research results interpretations of the term “computer software”**
Source: **compiled by the authors**

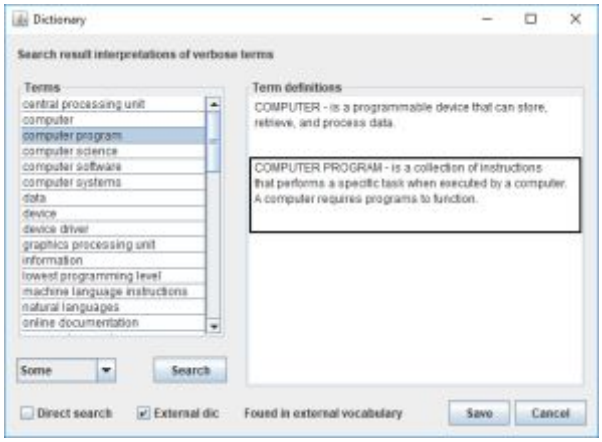

Fig. 4. **Research results for interpretations of the term “computer program”**

Source: **compiled by the authors**

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

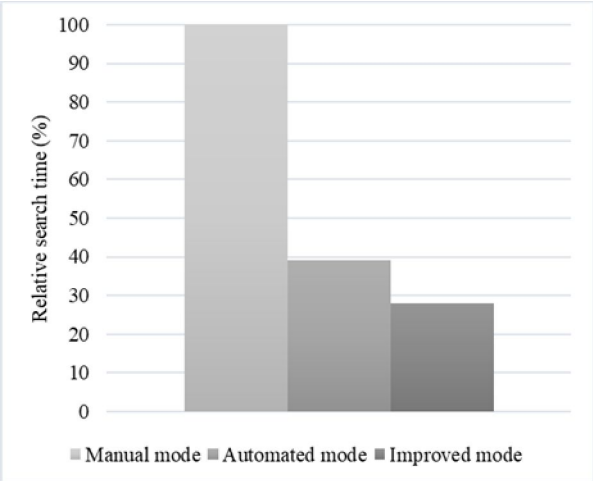

Fig. 5. **Time to determine the interpretation of terms in different modes**
Source: **compiled by the authors**

VII. CONCLUSIONS

The analysis of the existing methods of constructing DD for software projects and the need to reduce the time for the preparation of descriptions of terms are made. The method of definition of interpretations is proposed. This method allows to automate the process of determining the definitions of terms of subject areas. Algorithms and software that implement the proposed method is developed. The experiments have confirmed the effectiveness of the proposed solutions. The results of the study can be used in the development of software products "by order" at various stages of design: collection and analysis of requirements, database development, writing project documentation, etc.

REFERENCES

1. Kungurcev, A. B., Potochnyak, Ya. V. & Silyaev, D. A. “Method of automated construction of explanatory dictionary of subject area”. *Technology audit and production reserves.* 2015; Vol. 2 Issue 2 (22): 58–63. DOI: https://doi.org/10.15587/2312-8372.2015.40895.

2. Liubchenko V. & Sulimova, I. “Examining the attributes of transitions between team roles in the software development project”. *Eastern-European Journal of Enterprise Technologies.* 2017; 1/3 (85): 12–17. DOI: https://doi.org/10.15587/1729-4061.2017.91597.

3. “Best Practices for Data Dictionary Definitions and Usage Version 1.1.” Available from: https://www.pnamp.org/sites/default/files/best_practices_for_data_dictionary_definitions_and_usage_version_1.1_2006-11-14.pdf.

4. Califf, M. E. & Mooney, R. J. “Bottom-up relational learning of pattern matching rules for information extraction”. *Journal of Machine Learning Research.* 2003; 4: 177–210.

5. Hartmann, R. R. K. “Lexicography: Dictionaries, compilers, critics, and users. Routledge”. 2003. 412 p.

6. “Ways Data Dictionary Increases Software Developers Productivity”. Available from: https://dataedo.com/blog/ways-data-dictionary-increases-software-developers-productivity. 2018.

7. Bessmertny, I. A., Nugumanova, A. B., Mansurova, M. Ye. & Baiburin, Ye. M. “Method of rare term contrastive extraction from natural language texts”. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics.* 2017; 17(1): 81-91. DOI: https://doi.org/10.17586/2226-1494-2017-17-1-81-91.

8. Dal, V. I. “Explanatory Dictionary of Russian language. Part 1. Russian”. 2015. 636 p.

9. Ozhegov, S. I. Tolkovyy slovar' russkogo yazyka, [Dictionary of the Russian language] (in Russian). *Obrazovaniye, Oniks*, Moscow: Russian Federation. 2011. 736 p.

10. “Tolkovyy slovar' russkogo yazyka”. [Dictionary of the Russian language] (in Russian). – Available from: http://vedu.ru/expdic. – Active link: 8.10.2018.

11. “Oxford Dictionaries Online, 2018, Oxford University Press”. – Available from: https://www.oxforddictionaries.com/. – Active link: 18.10.2018.

12. “Slovar' na sayte Instituta yazykoznaniya im. A. A. Potebni”. [The dictionary on the website of the Institute of Linguistics. A. A. Potebni] (in Ukrainian). – Available from: http://www.inmo.org.ua/sum.html. – Active link : 9.10.2018.

13. Hasan, K. “Automatic key phrase extraction: a survey of the state of the art”. *Proc. 52nd Annual Meeting of the Association for Computational Linguistics.* 2014; No.1: 1262–1273. DOI: https://doi.org/10.3115/v1/p14-1119.

14. A. Melnikov et al. “On usage of machine learning for natural language processing tasks as illustrated by educational content mining”. *Ontology of designing.* 2017; Vol.7(1): 34–47. DOI: https://doi.org/10.18287/2223-9537-2017-7-1-34-47.

15. Braslavskii, P. & Sokolov, E. “Comparison of four methods for automatic two-word term extraction. Computational Linguistics and Intellectual Technology”. *Proc. Int. Conf. Dialog.* Moscow: Russian Federation. 2006. p. 88–94.

16. Vavilenkova, A. “Methods of identifying logical connections between parts of text documents”. Bulletin of the National Technical University “KhPI”. Series: New solutions in modern technologies. 2017; Issue 7(1229):118–122. DOI: https://doi.org/ 10.20998/2413-4295.2017.07.16.

17. Chainikova, G. R. “Development of learner’s electronic dictionary of a thesaurus type as means of forming internal foreign-language lexicon”. Bulletin of the Perm National Research Polytechnic University. 2014. p. 70–84.

18. “Institut programmnykh sistem RAN. Issledovatel'skiy tsentr iskusstvennogo intellekta, slovar' prostykh slovosochetaniy”. [Institute of Software Systems RAS. Artificial Intelligence Research Center, a dictionary of simple phrases] (in Russian). – Available from: http://ai-center.botik.ru/Airec/index.php/ru/resources/dictionaries/36-ru-collocation-dict – Active link : 12.11.2018.

19. Ushakov, D. N. “Tolkovyy slovar' sovremennogo russkogo yazyka”. [Explanatory Dictionary of Modern Russian] (in Russian). *Izdatel'stvo: Al'ta-print;* DOM XXI Vek. Moskva: Russian Federation. 2009. 1239 p.

20. Kungurtsev, O., Kovalchuk, S., Potochniak, Ia. & Shirokostup, M. “Creating the domain vocabulary on the basis of automated analysis of Ukrainian texts”. *Technical Sciences and Technologies.* 2016; No.3 (5): 164–174.

21. Kungurtsev, O., Zinovatnaya, S., Potochniak, Ia. & Kutasevych, M. “Development of information technology of term extraction from documents in natural language”. *Eastern-European Journal of Enterprise Technologies.* 2018; Vol.6 No.2 (96): 44–51. DOI: https://doi.org/10.15587/1729-4061.2018.147978.

“Slovnyk ukrayins′koyi movy”. [Dictionary of the Ukrainian language] (in Ukrainian). – Available from: http:// http://sum.in.ua/. – Active link: 20.10.2018.

22. Nielsen, Sandro. “The Effect of Lexicographical Information Costs on Dictionary Making and Use”. University of Aarhus. Lexicons. 2008; 18 (1): 170–189.

23. Navigli, R. & Velardi, P. “From Glossaries to Ontologies: Extracting Semantic Structure from Textual Definitions”. *IOS Press.* 2008. p.71–87.

24. John P. Considine. “Dictionaries in Early Modern Europe: Lexicography and the Making of Heritage”. *Cambridge University Press.* 2009; Vol.19: 509–515. DOI: https://doi.org/10.5788/19-0-452.

25. Bol'shakova, Ye. I., Klyshinskiy, E. S., Lande, D. V., Noskov, A. A., Peskova, O. V. & Yagunova Ye. V. Avtomaticheskaya obrabotka tekstov na yestestvennom yazyke i komp'yuternaya lingvistika, [Automatic natural language processing and computational linguistics] (in Russian). *MIEM.* Moscow: Russian Federation. 2011. 272 p.

26. Joanna, Olechno-Wasiluk. “Struktura slovarnoy stat'i v slovare “Rossiya. Bol'shoy lingvostranovedcheskiy slovar”. [The structure of an article entry in the dictionary Russian] (in Poland). Yezhegodnik Russko-pol'skogo Instituta. 2015. No.1 (8).

27. “Cognitive Technologies. Software package of syntactic analysis and machine translation “Cognitive Dwarf””. – Available from: http://www.cognitive.ru. – Active link: 12.09.2018.

28. “Interpretatio”. – Available from: http://www.softholm.com/download-software-free16427.htm. – Active link: 2.10.2018.

Conflicts of Interest: the authors declare no conflict of interest

Received 25.11.2018
Received after revision 05.02.2019
Accepted 21.02. 2019

DOI: https://doi.org/10.15276/aait.02.2019.1
УДК 004.912

МЕТОД ПОШУКУ ТЕРМІНІВ ДЛЯ СЛОВНИКІВ ПРЕДМЕТНИХ ОБЛАСТЕЙ, ЩО ВИКОРИСТОВУЮТЬСЯ ПРИ ПРОЕКТУВАННІ ПРОГРАМНИХ ПРОДУКТІВ

Олексій Борисович Кунгурцев[1)]
ORCID: https://orcid.org/0000-0002-3207-7315, abkun@te.net.ua
Світлана Леонідівна Зіноватна[1)]
ORCID: https://orcid.org/0000-0002-9190-6486, zinovatnaya.svetlana@opu.ua
Яна Володимирівна Поточняк[1)]
ORCID: https://orcid.org/ 0000-0003-1291-1146, yana.onpu@gmail.com
[1)] Одеський національний політехнічний університет, пр. Шевченко, 1. Одеса, 65044, Україна

АНОТАЦІЯ

В роботі запропоновано спосіб формування тлумачень термінів для словника предметної області з використанням існуючих тлумачних словників. Запропоновано математичний опис процедур пошуку тлумачень і алгоритмів, які їх реалізують. Запропоновано метод автоматизованого пошуку тлумачення термінів, який характеризується підключенням існуючих словників, оцінкою якості тлумачення за наявністю в ньому термінів з предметної області, синтезом тлумачення в разі розбіжності термінів, що дозволяє скоротити час роботи експерта зі словником. Розроблено програмне забезпечення, що дозволяє істотно прискорити процес пошуку тлумачень термінів для словника предметної області. В роботі запропонований спосіб формування визначень термінів для словника предметної області з використанням існуючих тлумачних словників. Показано, що при об'єднаному пошуку термінів і їх тлумачень можна знайти близько 10 % дефініцій, що явно недостатньо. Запропоновано метод автоматизованого пошуку тлумачення термінів, що передбачає використання існуючих тлумачних словників. Запропоновано математичну модель словникової статті словника предметної області. Запропоновано математичну модель словникової статті тлумачного словника, що враховує найголовніше слово, множину тлумачень слова, послід і стійкі словосполучення. Розроблено механізм виділення дефініцій терміна з тлумачного словника в залежності від структури його словникової статті. Розроблено алгоритм автоматизованого пошуку дефініцій для однослівних термінів. Розроблено алгоритм автоматизованого пошуку дефініцій для багатослівних термінів, заснований на виділенні з терміна іменників. Запропоновано механізм оцінки якості можливих тлумачень в залежності від входження в них термінів з предметної області терміна. Розроблено механізм вибору дефініцій при неповному збігу термінів зі словника предметної області та тлумачного словника, заснований на процедурі декомпозиції терміна, пошуку часткових тлумачень і синтезу результуючого тлумачення. Розроблено програмне забезпечення, яке дозволяє організувати пошук тлумачень термінів як в локальних тлумачних словниках (раніше завантажених в систему), так і в онлайн-словниках. У завдання експерта входить оцінка знайдених тлумачень і можливо, їх редагування. Експериментальна оцінка ефективності застосування програмного продукту показала скорочення часу роботи експерта порівняно з «ручним режимом» приблизно в 4 рази.

Ключові слова: словник предметної області; однослівний термін; багатослівний термін; морфологічний аналіз; математична модель терміна; тлумачення терміна; текстовий документ

DOI: https://doi.org/10.15276/aait.02.2019.1
УДК 004.912

МЕТОД ПОИСКА ТОЛКОВАНИЙ ТЕРМИНОВ ДЛЯ СЛОВАРЕЙ ПРЕДМЕТНЫХ ОБЛАСТЕЙ, ИСПОЛЬЗУЕМЫХ ПРИ ПРОЕКТИРОВАНИИ ПРОГРАММНЫХ ПРОДУКТОВ

Алексей Борисович Кунгурцев[1)]
ORCID: https://orcid.org/ 0000-0002-3207-7315, abkun@te.net.ua
Светлана Леонидовна Зиноватная[1)]
ORCID: https://orcid.org/0000-0002-9190-6486, zinovatnaya.svetlana@opu.ua
Яна Владимировна Поточняк[1)]
ORCID: https://orcid.org/ 0000-0003-1291-1146, yana.onpu@gmail.com
[1)]Одесский национальный политехнический университет, пр. Шевченко, 1. Одесса, 65044, Украина

АННОТАЦИЯ

В работе предложен способ формирования определений терминов для словаря предметной области с использованием существующих толковых словарей. Предложено математическое описание процедур поиска толкований и алгоритмы, которые их реализуют. Предложен метод автоматизированного поиска толкования терминов, который характеризуется подключением существующих словарей, оценкой качества толкования по наличию в нем терминов из предметной области, синтезом толкования в случае несовпадения терминов, что позволяет сократить время работы эксперта со словарем. Разработано программное обеспечение, позволяющее существенно ускорить процесс поиска толкований терминов для словаря предметной области.

Ключевые слова: словарь предметной области; однословный термин; многословный термин; морфологический анализ; математическая модель термина; толкование термина; текстовый документ

ABOUT THE AUTHORS

Oleksii B. Kungurtsev, PhD (Eng), Professor of the Department of System Software, Odessa National Polytechnic University, 1, Shevchenko Avenue. Odesa, 65044, Ukraine
abkun@te.net.ua. ORCID: https://orcid.org/0000-0002-3207-7315.

Олексій Борисович Кунгурцев, канд. техніч. наук, професор каф. Системного програмного забезпечення, Одеський національний політехнічний університет, пр. Шевченко, 1. Одеса, 65044, Україна

Svetlana L. Zinovatnaya, PhD (Eng), Associate Professor of the Department of System Software, Odessa National Polytechnic University, 1, Shevchenko Avenue. Odesa, 65044, Ukraine
zinovatnaya.svetlana@opu.ua. ORCID: https://orcid.org/ 0000-0002-9190-6486.

Світлана Леонідівна Зіноватна, кандидат техніч. наук, доцент каф. Системного програмного забезпечення. Одеський національний політехнічний університет, пр. Шевченко, 1. Одеса, 65044, Україна

Iana V. Potochniak, Postgraduate Student of the Department of System Software, Odessa National Polytechnic University, 1, Shevchenko Avenue. Odesa, Ukraine, 65044
yana.onpu@gmail.com. ORCID: https://orcid.org/ 0000-0003-1291-1146.

Яна Володимирівна Поточняк, аспірантка аспірантка каф. Системного програмного забезпечення.
Одеський національний політехнічний університет, пр. Шевченко, 1. Одеса, 65044, Україна