Ways of spinning implementation in complex natural language sentences

Anastasiia Vavilenkova and Sergiy Gnatyuk

National Aviation University, Liubomyr Huzar Ave.,1, Kyiv, 03058, Ukraine

Abstract

This article is dedicated to different ways of searching for rewritten sentences in textual information. It is possible to make this research by means of logic and linguistic modeling. Each natural language sentence can be transformed into various types of logic and linguistic knowledge representation models. Every type of spinning implementation can be corresponded with special type of logic and linguistic model. The author demonstrated particular methods of spinning recognizing, that are based on the logic of predicates.

Keywords

Natural language, logic and linguistic models, text information, spinning, rewrite

1. Introduction

Nowadays the most critical problem of our society is finding new content for a significant number of sites, articles from different areas, and new significant research in scientific materials. Unfortunately, the statistic of plagiarism in Ukraine as in other countries in the world is very sad. For instance, a national survey published in Education Week found that 54% of students in the USA admitted to plagiarizing from the Internet and 74% admitted that at least once during the past school year they had engaged in "serious" cheating [12-13]. In 2016 according to the polish investigations Ukraine engaged the fifth position in academic plagiarism among the students [14]. Plagiarism statistics among the students in National Aviation University on technic specialties is no exception (Figure 1).

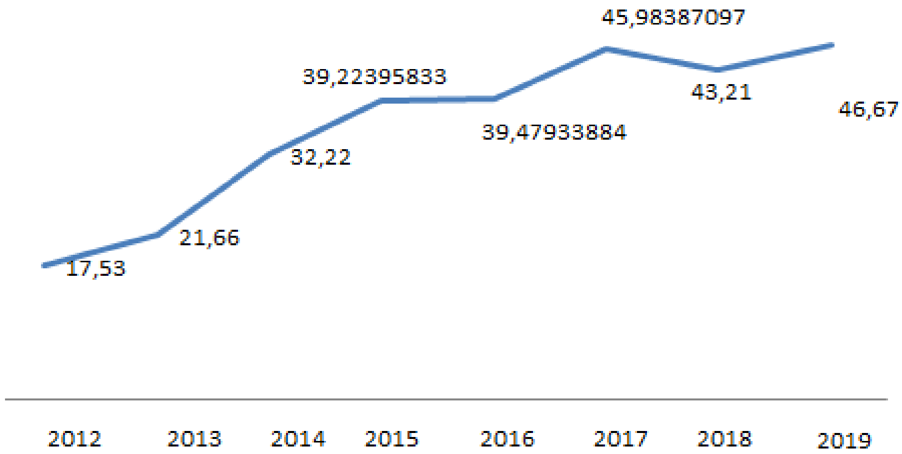

Figure 1: Plagiarism statistic among the students of technic specialties

CMiGIN 2022: 2nd International Conference on Conflict Management in Global Information Networks, November 30, 2022, Kyiv, Ukraine
EMAIL: vavilenkovaa@gmail.com (A. Vavilenkova); s.gnatyuk@nau.edu.ua (S. Gnatyuk)
ORCID: 0000-0002-9630-4951 (A. Vavilenkova); 0000-0003-4992-0564 (S. Gnatyuk)

It is distinguished such types of plagiarism [15]:

- Rewriting a text from the source by your own without referencing;
- Downloading essays, articles and other types of works form the open sites like yours without referencing;
- Reported that they copied verbatim from written sources without any references;
- Copywriting somebody texts with changing the worlds' order without referencing;
- Using somebody texts verbatim and referencing for another source;
- Translation from another language without referencing;
- Referencing for your own academic works;
- Admitted to writing false and fabricated bibliography records;
- Group work without author participation.

On Figure 2 we can see diagram with the outstanding examples of plagiarism.

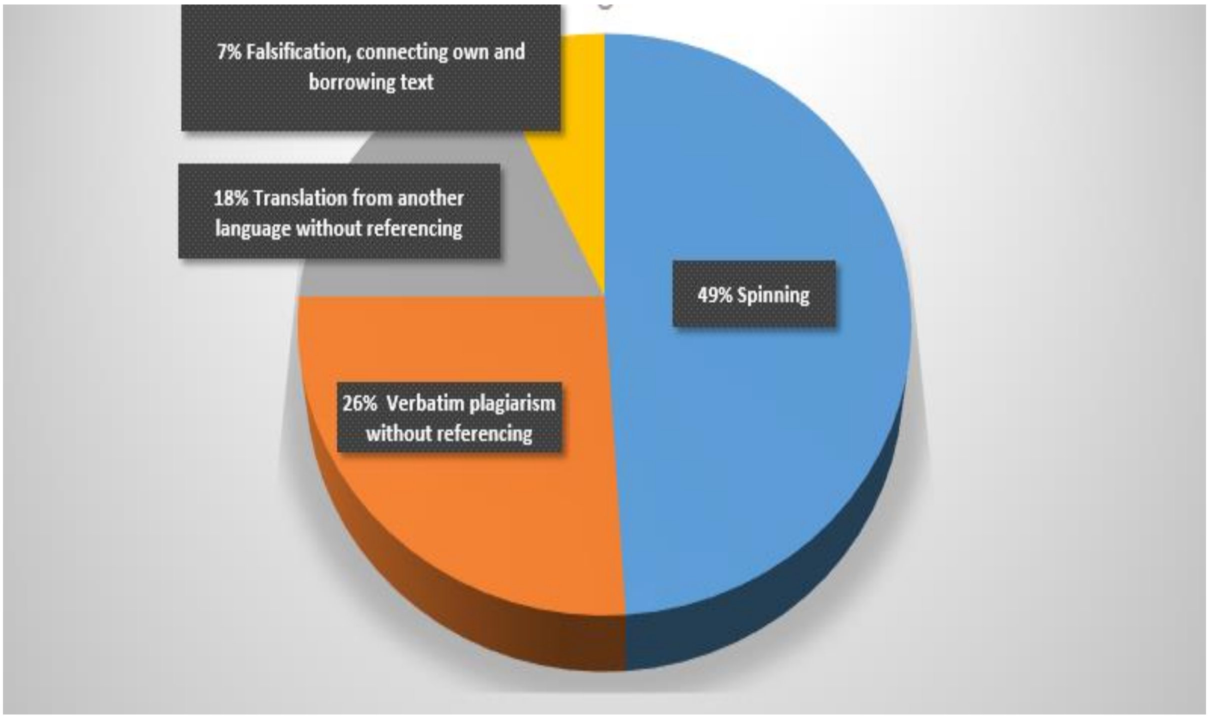

Figure 2: Diagram with the outstanding examples of plagiarism

Plagiarism, which is to say duplicate, copy and rewrite textual information into the Internet, leads to reduction of student's knowledge and author's rights contravention. According to requirements of shape periodicals it is possible to use less than 30% of yourself references [1]. Appearance of new technologies make the process of electronic textual information copywriting easier. Different types of computer modelling falsifications, edition of graphics, video and audio materials also raise possibility of plagiarism [16].

Today in 2020s almost all teachers are sharing their lecture and practice materials by mean of various resources, so cheating among the students and pupils will be more than ever before [17], for example, during the writing control works.

There are various programs and additions for solving this problem [18-21], but what is their quality?

Analysis and testing of outcomes of these systems give opportunity to detect a set of functions, that they are not implemented, for instance, calculation the present of coincidences with analysis of references, comparing the context of textual information.

Thus, the aim of this article is solving the problem of context comparing of textual information. The author created the rules for searching for rewrite sentences in electronic textual documents by mean of logic and linguistic modelling.

2. Materials and methods

A lot of ways of spinning implementation are based on technical factors, for instance:

- Changing some symbols, that are similar for different flexional languages ("i", "o", ect.);
- Incorrect referencing;
- Absent of one legislative system for comparing electronic textual information with unique measures for estimation;
- Calculation the present of coincidences with analysis of self-references and other references.

However, all these lacks of present systems for comparing textual information, can be correct by the programmers. Solving the problem of context identity of spinning implementation needs usage of special knowledge from computer linguistic [2, 5] and linguistic analysis [6-7].

As an outcome of various ways of formal interpretation of synonymic constructions into the sentences of natural languages [8], it had been created various types of logic and linguistic knowledge representation models, that can be corresponded with each type of spinning context implementation.

Each natural language sentence can be transformed into the various types of logic and linguistic knowledge representation models [3, 11].

Invariant logic and linguistic model Q^S is corresponded to the same natural language sentence S by the context, witch interprets by mean of logic and linguistic model L^S. Both of the models are simple predicates.

Let original natural language sentence S is depicted by logic and linguistic model L^S:

$$L^S = p_1(x_1, g_1, y_1, q_1, z_1, r_1, h_1)$$

and invariant logic and linguistic model Q^S is

$$Q^S = p_2(x_2, g_2, y_2, q_2, z_2, r_2, h_2).$$

These are the base semantic models of conversional derivation.

Spinning implementation 1. If the subject of logic and linguistic model x_1 is similar for object y_2 in instrumental case and predicate p_1 is the verb, that is typical for noun x_1 (they are collocation into the knowledge base), the object y_1 is the same as sub-matter of relation z_2, then the models L^S and Q^S have identical context.

For example, natural language sentence "*Pupils quietly set in a large classroom*" has such a logic and linguistic model:

$$L^S = p_1(x_1, 0, y_1, q_1, 0, 0, h_1),$$

$L^S =$ set (pupils, 0, classroom, large, 0,0, quietly).

Logic and linguistic model for another natural language sentence "*Presented people were pupils in the classroom*" will be:

$$Q^S = p_2(x_2, g_2, y_2, 0, z_2, 0, 0),$$

$Q^S =$ were (people, presented, pupils, 0, classroom,0,0).

According to the rule, the context of these sentences is similar and the model Q^S is invariant to L^S.

Spinning implementation 2. If subjects of both logic and linguistic models are similar $x_1 \equiv x_2$, predicates p_1 and p_2 are from the same time and predicate p_2 have one root with object y_1 or sub-matter z_1, and $y_1 \equiv y_2$ or $z_1 \equiv y_2$, then the models L^S and Q^S have identical context.

For instance, natural language sentence "*The girl went to the mountain top with the skis*" can be depict with logic and linguistic model:

$$L^S = p_1(x_1, 0, y_1, q_1, z_1, 0, 0),$$

$L^S =$ went (girl, 0, top, mountain, skis,0, 0).

Logic and linguistic model for another natural language sentence "*The girl was skiing from the mountain top*" will be:

$$Q^S = p_2(x_2, 0, y_2, q_2, 0,0,0),$$

$$Q^S = \text{was\_skiing (girl, 0, top, mountain, 0,0,0).}$$

According to the spinning implementation 2, the context of these two sentences is similar and the model Q^S is invariant to L^S.

For the next sentences "*The man lived into the building near the park*" and "*The man had apartment next the park*" we will have such models:

$$L^S = p_1(x_1, 0, y_1, 0, z_1, 0, h_1),$$

$$L^S = \text{lived (man, 0, building, 0, park,0, near).}$$

$$Q^S = p_2(x_2, 0, y_2, 0, z_2, 0, h_2),$$

$$Q^S = \text{had (man, 0, apartment, 0, park,0, next).}$$

The sentences have the same structure, so they have the same type of logic and linguistic model, in witch were used synonyms "*building*" and "*apartment*", "*near*" and "*next*". However the words "*lived*" and "*had*" are not synonyms, they are conversions, so the context of these two sentences is similar and the model Q^S is invariant to L^S.

Spinning implementation 3. If objects of the first sentence y_1 or it's characteristic is similar to subject of the second sentence $y_1 \equiv x_2$ or $q_1 \equiv x_2$, and $z_1 \equiv y_2$, $r_1 \equiv q_2$, and predicate of the second sentence has one root with y_1 or z_1, then the models L^S and Q^S have identical context.

Natural language sentence "*Some boy carried the girl's ski on the mountain top*" interpret by mean of logic and linguistic model:

$$L^S = p_1(x_1, g_1, y_1, q_1, z_1, r_1, 0),$$

$$L^S = \text{carried (boy, some, ski, girl's, top, mountain, 0).}$$

Logic and linguistic model for another natural language sentence "*The girl was skiing from the mountain top*" will be:

$$Q^S = p_2(x_2, 0, y_2, q_2, 0,0,0),$$

$$Q^S = \text{was\_skiing (girl, 0, top, mountain, 0,0,0).}$$

According to the spinning implementation 3, the contexts of these two sentences are similar and the model Q^S is invariant to L^S.

The next spinning implementations connect with changing the part of speech for the world of natural language without changing lexical meaning. The order of sentences, that compare, does not matter.

Spinning implementation 4. If subjects of the both sentence are identity $x_1 \equiv x_2$, predicates of the sentences are synonyms, object and sub-matter of the sentences are equal $z_1 \equiv y_2$ or $z_2 \equiv y_1$, then the models L^S and Q^S have identical context.

Natural language sentence "*Traditions begin from the history*" interpret by mean of logic and linguistic model:

$$L^S = p_1(x_1, 0, y_1, 0,0,0,0),$$

$$L^S = \text{begin (traditions, 0, history, 0, 0, 0, 0).}$$

Logic and linguistic model for another natural language sentence "*Traditions take their roots into the history*" will be:

$$Q^S = p_2(x_2, 0, y_2, q_2, z_2, 0,0),$$

$$Q^S = \text{take (traditions, 0, roots, their, history,0,0).}$$

According to the spinning implementation 4, the contexts of these two sentences are similar and the model Q^S is invariant to L^S.

Spinning implementation 5. If predicates of both sentences are different grammar forms of the one word $p_1 \equiv p_2$, $p_1 \in P^S$, $p_2 \equiv P^S$ and $x_1 \equiv z_2$, $x_1 \in X_p^S(h)$, $z_2 \in Z_p^S(x,g,y,q,h)$, $y_1 \equiv y_2$, $y_1 \in Y_p^S(x,g,h)$, $y_2 \in Y_p^S(x,g,h)$, then the models L^S and Q^S have identical context.

For the next sentences "*This book is very interesting for boys*" and "*The boys are very interested in this book*" we will have such models:

$$L^S = p_1(x_1, g_1, y_1, 0,0,0,0),$$

$L^S =$ interesting (book, this, boys, 0, 0,0, 0).

$$Q^S = p_2(x_2, 0, y_2, q_2, 0,0,0),$$

$Q^S =$ interested (boys, 0, book, this, 0,0, 0).

Spinning implementation 6. If predicates of both sentences are different grammar forms of the one word $p_1 \equiv p_2$, $p_1 \in P^S$, $p_2 \equiv P^S$, $x_1 \equiv y_2$, $y_1 \equiv x_2$, $x_1 \in X_p^S(h)$, $y_2 \in Y_p^S(x,g,h)$, $y_1 \in Y_p^S(x,g,h)$, $x_2 \in X_p^S(h)$, $z_1 \equiv z_2$, $z_1 \in Z_p^S(x,g,y,q,h)$, $z_2 \in Z_p^S(x,g,y,q,h)$, so we have one sentence in active and another sentence in passive, then the models L^S and Q^S have identical context.

For the next sentences "*The scientists used new methods in their practice*" and "*New methods are used in practice by scientists*" we will have such models:

$$L^S = p_1(x_1, 0, y_1, q_1, z_1, r_1, 0),$$

$L^S =$ used (scientists, 0, methods, new, practice, their, 0).

$$Q^S = p_2(x_2, g_2, y_2, 0, z_2, r_2, 0),$$

$Q^S =$ used (methods, new, scientists, 0, practice, their, 0).

Spinning implementation 7. If predicates of both sentences are similar $p_1 \equiv p_2$, $p_1 \in P^S$, $p_2 \equiv P^S$, $x_1 \equiv y_2$, $x_1 \in X_p^S(h)$, $y_2 \in Y_p^S(x,g,h)$, $y_1 \equiv x_2$, $y_1 \in Y_p^S(x,g,h)$, $x_2 \in X_p^S(h)$, then the models L^S and Q^S have identical context.

For the next sentences "*The magnet is gravitates to the iron*" and "*The iron is gravitates to the magnet*" we will have such models:

$$L^S = p_1(x_1, 0, y_1, 0,0,0,0),$$

$L^S =$ gravitates (magnet, 0, iron, 0, 0, 0, 0).

$$Q^S = p_2(x_2, 0, y_2, 0,0,0,0),$$

$Q^S =$ gravitates (iron, 0, magnet, 0, 0, 0, 0).

According to the spinning implementation 7, the context of these sentences is similar and the model Q^S is invariant to L^S.

Spinning implementation 8. If predicates of both sentences are similar $p_1 \equiv p_2$, $p_1 \in P^S$, $p_2 \equiv P^S$, $y_1 \equiv z_2$, $y_1 \in Y_p^S(x,g,h)$, $z_2 \in Z_p^S(x,g,y,q,h)$, $z_1 \equiv y_2$, $z_1 \in Z_p^S(x,g,y,q,h)$, $y_2 \in Y_p^S(x,g,h)$, then the models L^S and Q^S have identical context.

For example, natural language sentence "*Scientists developed new methods for different countries*" has such a logic and linguistic model:

$$L^S = p_1(x_1, 0, y_1, q_1, z_1, r_1, 0),$$

$L^S =$ developed (scientists, 0, methods, new, countries, different, 0).

Logic and linguistic model for another natural language sentence "*Scientists developed new methods to the different countries*" will be:

$$Q^S = p_2(x_2, 0, y_2, q_2, z_2, r_2, 0),$$

$Q^S =$ developed (scientists, 0, methods, new, countries, different, 0).

According to the rule, the context of these sentences is similar and the model Q^S is invariant to L^S.

Spinning implementation 9. If predicates of both sentences are different grammar forms of the one word $x_1 \equiv y_2$, $x_1 \in X_p^S(h)$, $y_2 \in Y_p^S(x,g,h)$, $y_1 \equiv x_2$, $y_1 \in Y_p^S(x,g,h)$, $x_2 \in X_p^S(h)$, so we have one sentence in active and another sentence in passive, then the models L^S and Q^S have identical context.

For example, natural language sentence "*The boy gives the girl a present*" has such a logic and linguistic model:

$$L^S = p_1(x_1,0,y_1,0,z_1,0,0),$$

$L^S =$ gives (boy, 0, girl, 0, present, 0, 0).

Logic and linguistic model for another natural language sentence "*The girl takes a present from the boy*" will be:

$$Q^S = p_2(x_2,0,y_2,0,z_2,0,0),$$

$Q^S =$ takes (girl, 0, boy, 0, present, 0, 0).

According to the rule, the context of these sentences is similar and the model Q^S is invariant to L^S.

Spinning implementation 10. If one sentence of natural language is simple and describes by logic and linguistic model

$$L^S = p_1(x_1,g_1,y_1,q_1,z_1,r_1,h_1)$$

and another sentence on natural language is complex with logic operation of implication:

$$Q^S = p_2(x_2,g_2,y_2,q_2,z_2,r_2,h_2) \rightarrow p_2'(x_2',g_2',y_2',q_2',z_2',r_2',h_2'),$$

that interprets conditions, where $y_1 \equiv y_2$, $z_1 \equiv z_2$, h_1 and h_2 are antonyms, $x_1 \equiv x_2'$, $g_1 \equiv g_2'$, $x_2 \equiv 0$, y_1 and p_2' have one root, then the models L^S and Q^S have identical context.

For example, natural language sentence "*The scientists often used new methods in their practice*" has such a logic and linguistic model:

$$L^S = p_1(x_1,0,y_1,q_1,z_1,r_1,h_1),$$

$L^S =$ used (scientists, 0, methods, new, practice, their, often).

Logic and linguistic model for another natural language sentence "*If something often used in scientists practice, so that are new methods*" will be:

$$Q^S = p_2(x_2,g_2,y_2,q_2,z_2,r_2,h_2) \rightarrow p_2'(x_2',g_2',y_2',q_2',z_2',r_2',h_2'),$$

$Q^S =$ used (something, 0, practice, scientists, 0, 0, often) $\rightarrow$
used (scientists, 0, methods, new, 0, 0, often).

According to the spinning implementation 10, the context of these sentences is similar and the model Q^S is invariant to L^S.

3. Experiment

The study proposes the rules for spinning implementation detection. They were used into the system of comparing analysis of electronic text documents. This system also based on the basic principles and rules for the synthesis of logic and linguistic models of natural language sentences and abstract models of logical conversion, that have been created to formalize the description of logical relationships between parts of text documents and their geometric interpretations.

All these facts give opportunity to compare results of work for modern systems of comparing analysis and the system, proposed by author.

It was made experience under the test textual information according to the time of verification and percent of conjunction. In the Table 1, Table 2 and Table 3 we can see different indicates of comparing analysis, made by different systems for the same texts.

On the said figures were demonstrated average results of the time for the comparing process and the average percent of conjunction, because of impossibility of checking big text at all. All modern systems proposed limited amount of worlds for free comparing. Thus, for experiment it had been needed to divide text for more less parts and had been analyzed real time for processing.

As we see, proposed system of automated analysis loses in time of processing, but wins in percent of conjunction, that approved possibility of usage of supported rules for spinning implementation detection.

Table 1

Results of experiments by systems Advego Plagiatus and Text.ru

Type of work	Amount of symbols	Amount of words without spaces	Amount of words	Advego Plagiatus Time	Advego Plagiatus %	Text.ru Time	Text.ru %
Tethis 1	6126	5290	868	4 min. 30s.	72	1 min.	96.26
Tethis 2	6648	5817	807	1 min. 39s.	64	1 min.	90.86
Diploma of baccalaure	67597	58951	10809	10 min. 20s.	15	3min.	15.57
ate	14188	12152	1917	3 min. 10s.	16	2min.	12.66
Article 1	33049	28726	5095	5 min. 30s.	70	2min.	72.48
Article 2	135190	129340	20655	11 min.	34	3 min.	36.52
Diploma of master	377385	326773	60081	15 min.	10	3 min.	9.56
Diploma of master degree							

Table 2

Results of experiments by systems Strike Plagiarism and System “Антиплагіат”

Type of work	Number of symbols	Amount of words without spaces	Amount of words	Strike Plagiarism Time	Strike Plagiarism %	System “Анти-плагіат” Time	System “Анти-плагіат” %
Tethis 1	6126	5290	868	2 min.	86,7	3 min. 50s.	74
Tethis 2	6648	5817	807	2 min.	80,5	3 min. 50s.	90
Diploma of baccalaureat	67597	58951	10809	15 min.	18,2	3 min. 50s.	21
e	14188	12152	1917	10 min.	16,1		56
Article 1	33049	28726	5095	10 min.	2,6	3 min.	45
Article 2	135190	129340	20655	15 min.	35,4	50s. 3	100

Type of work	Number of symbols	Amount of words without spaces	Amount of words	Strike Plagiarism Time	Strike Plagiarism %	System “Анти-плагіат” Time	System “Анти-плагіат” %
Diploma of master	377385	326773	60081	25 min.	5,7	min. 50s. 3 min. 50s.	17
Diploma of master degree						3 min. 50s.	

Table 3
Results of experiments by proposed system of comparing analysis

Type of work	Amount of symbols	Amount of words without spaces	Amount of words	Proposed system of comparing analysis Time	Proposed system of comparing analysis %
Tethis 1	6126	5290	868	5 min.	97,1
Tethis 2	6648	5817	807	5 min.	92,5
Diploma of baccalaureate	67597	58951	10809	12 min.	18,5
Article 1	14188	12152	1917	8 min.	60,2
Article 2	33049	28726	5095	8 min.	76,4
Diploma of master	135190	129340	20655	12 min.	100
Diploma of master degree	377385	326773	60081	15 min.	15,5

4. Conclusions

Due to the lack of adequate formal models of natural language objects and the fact that relevant problems require informal, creative human input, a computer is still unable to fully resolve the problem of text information despite nearly a century of artificial intelligence research [4].

We must create a system that enables quick surface structure analysis and the building of a reasonably straightforward and strict semantic configuration in order to explain the overall meaning of the text [9–10]. In order to create such a model, we must figure out how to get the specific objects and relationships that the text implicitly represented.

The significance of finding a solution to this issue in information retrieval systems is the requirement to focus the search, excluding documents that refer to the user's useless objects, and to safeguard against the possibility that the user may request an object using different words or phrases than the author uses to describe an event.

The key phase in the algorithm for creating a meaningful model of text is the synthesis of linguistic and logical models, which is based on construction principles and the lookout for fundamental relationships. The relationships mentioned above analyze natural language phrases with an arbitrary form and have equivalent substance.

In contrast to existing methods of searching for text duplicates, the study suggests rules for automatic determination of the logical identity of complex and simple predicates that are part of the logic and linguistic models of electronic text documents. These rules are based on content analysis, rules, and models of constructing complex synonymous designs, which improves the evaluation of the accuracy of the results.

References

[1] C. B. Carter, C. F. Blanford, Plagiarism and detection, Journal of Materials Science 51(2016) 7047–7048. https://doi.org/10.1007/s10853-016-0004-7.

[2] W. Che, Y. Zhang, Deep learning in lexical analysis and parsing, Deep learning in Natural Language Processing, 2018. http:// doi.org/10.1007/978-981-10-5209-5_4.

[3] A. Vavilenkova, Basic principles of the synthesis of logical-linguistic models, Cybernetics and systems analysis 51(5) (2015) 826–834. http:// doi.org/10.1007/s10559-015-9776-z.

[4] P. Osipov, A. Borisov, Practice of Web Data Mining Methods Application, Scientific journal of Riga Technical University 5 (2010) 101–107. https://doi.org/10.2478/v10143-010-0014-x.

[5] V. Evans, Lexical concepts, cognitive models and meaning-construction, Cognitive Linguistic 17 (2006) 73–107.

[6] S. Th. Gries, A. Stefanovich (Eds.), Corpora in Cognitive Linguistics. Corpus-Based Approaches to Syntax and Lexis, in: Trends in Linguistics, Berlin/New York, Mouton de Gruyter, 2006.

[7] A. Vavilenkova, Analysis and synthesis of logic and linguistic models, TOV “SIK GROUP Ukraine”, 2017.

[8] I. R. Galperin, Text as an object of linguistic research, 5th. ed., KomKniga, 2007.

[9] I. M. Kobozeva, Linguistic semantic, M., Editorial, 2000.

[10] D. Geeraerts, R. Dirven and R. John, Cognitive linguistics: basic readings research, Publ. mouton de Cruyter, Berlin – New York, 2006.

[11] A. Vavilenkova, Modelling of the context links between the natural language sentences, in: Proceedings of the 9th International Scientific and Practical Conference "Information Control Systems & Technologies" (ICST2020), 2020, pp. 282–293.

[12] Cyber Plagiarism & Statistics, 2020. https://www.checkforplagiarism.net/cyber-plagiarism.

[13] L. Keegan, 79+Staggering Plagiarism Statistics You Need to Know, 2020. https://skillscouter.com/plagiarism-statistics/.

[14] The level of plagiarism into the students works in Europe, 2020. http://aphd.ua/riven-plahiatu-v-studentskykh-robotakh-u-kranakh-ievropy-/.

[15] An epidemic of academic plagiarism in digits, 2016. https://studway.com.ua/plagiat-2/.

[16] Academic Integrity in 2020, Plagiarim Today, 2020. https://www.plagiarismtoday.com/2020/08/19/academic-integrity-in-2020/.

[17] V. Baxrushin, Academic plagiarism and self-plagiarism in science and high education: normative base and world experience, 2018. http://education-ua.org/ua/articles/1128-akademichnij-plagiat-i-samoplagiat-v-nautsi-ta-vishchij-osviti-normativna-baza-i-svitovij-dosvid.

[18] Plagiarism Checker that Prefers Results over Numbers, 2020. https://unicheck.com/.

[19] On-line service for checking text for originality, 2020. https://text.ru/.

[20] StrikePlagiarism, 2020. http://strikeplagiarism.com.

[21] Checking of text originality Advego Plagiatus, 2020. http://advego.ru.