ЗБІРНИК НАУКОВИХ ПРАЦЬ ◊ СБОРНИК НАУЧНЫХ ТРУДОВ ◊ COLLECTION OF SCIENTIFIC PAPERS

ΛΌΓΟΣ

№2

Materials of International Scientific and Practicial Conference
«Scientific Development and Achievements»
Desember 1, 2017; St. Andrews, Scotland, UK

MATERIALS OF INTERNATIONAL SCIENTIFIC AND PRACTICIAL CONFERENCE

SCIENTIFIC DEVELOPMENT AND ACHIEVEMENTS

1 DECEMBER 2017

St Andrews, Scotland, United Kingdom

PART 2

S 30 **Scienti ic development and achievements** : materials of International scientific and practical conference December 1, 2017 in St. Andrews, Scotland / ed. for the production M.A. Holdenblat. Obuhiv : Printing House «Drukarik», 2017. Part. 2. P. 230.

Presents the article abstracts and participants of the international scientific-practical conference «Scientific development and achievements» held in St. Andrews, Scotland, UK, December 1, 2017.

The journal is dedicated to students, graduates, postgraduates, doctoral candidates, PhDs, young specialists, lecturers, scientists and others interested persons, as well as for a wide range of readers.

Bibliographic description of the materials of the conference is presented in the scientific electronic library «Elibrary.ru».

The collection of scientific papers is included to the international science and technology databases RINC and Google Academy.

CONTENT

SECTION 11.
ECONOMICS

CONTENT

CONTENT

SECTION 12.
TECHNICAL SCIENCES

CONTENT

CONTENT

CONTENT

UDC 003.26

THE WAYS OF SOLVING PROBLEMS OF DATA PROCESSING AUTOMATION IN THE SYSTEMS OF TEXTUAL STEGANALYSIS

Iryna Fedotova-Piven
Ph.D., Assistant Professor of the Department of Information Security and Computer Engineering
Cherkassy state technological university
Yaroslav Tarasenko
Postgraduate student of the department of information security and computer engineering
Cherkassy state technological university
Ukraine

Summary. The article deals with the problems that arise in the process of the textual data processing automation. It is proved that the basic methods of steganography are constructed on the basis of vulnerabilities caused by the problems of the textual analysis automation. An overview of the existing ways of solving these problems in morphological and syntactic text research programs is conducted. It is proposed to solve the problems of the data processing automation by creating a software complex which takes into account the features of the steganalysis for the needs of the systems of textual steganalysis.

Keywords: *automated text processing, textual steganalysis, automation problems, morphological analysis, syntactic analysis, textual semantic.*

Formulation of the problem. During the textual information processing automation for the creation of software tools and systems of text steganalysis arise some problems associated with the transferring of classical linguistic methods of text analysis to the computer research area. Identifying ways of solving these problems will contribute to the effective development of textual steganalysis systems.

Analysis of research and publications. V. Beloziorov in his theses on topic "Preparation the initial data for documents' clusterization" made a significant contribution to the selection and classification of problems that arise in the process of the text processing automation. Researches of such authors as Olha Babina, Nikita Diumin, Michaela Regneri, Diane King are intended to solve the problems of morphological, syntactic and discursive analysis automation. Ivan Nechta, Zhili Chen, including Ukrainian author Nataliia Kukharska conduct research of steganography, including textual steganography, as well as issues of linguistic steganalysis.

The purpose of the article. The aim of the work is to determine the existing and prospective ways of solving the problems of the textual data processing automation in the systems of textual steganalysis and to identify the possible ways of the text-research process automation for the development of software tools and systems of textual steganalysis.

Presenting main material. During the modern development of the society in the direction of informatization and the formation of an information society, due to the proliferation of computer processing and data transmission facilities appear threats related to the secret or commercial information leakage, illegal communication and cyberterrorism. As noted in [1], it is extremely necessary to develop automated computer systems for the analysis of the text [1, p. 51] which could effectively investigate the textual data and detect any traces of deliberately concealed information. It follows the need to automate the methods of linguistic research of the text, and on this path a number of problems arise. In particular, V. Beloziorov in his in his theses on topic "Preparation the initial data for documents' clusterization" [2] considers the difficulties that arise in the process of automated text processing. Thus, the author highlights the main problems of automated processing of the documents, namely the wrong spelling of the words in the text, different options for writing the same word, the developed morphology of some languages, the dependence on

the context, the presence of homonyms [2, p. 5]. All of these affect the effectiveness of automated text analysis in general and staganalysis in particular. Major attacks on the linguistic stegosystem are constructed on the basis of the mentioned problems of automated textual data processing. It follows the need to identify the ways to overcome these vulnerabilities. In [3], the author distinguishes three main groups of attacks on a linguistic stegosystem, namely methods of random intervals, syntactic and semantic methods. Each of these groups of methods is based on the relevant vulnerabilities caused by the difficulties of automatic text processing. For example, syntactic methods can be compared with vulnerability based on context dependence, semantic with the presence of homonyms and the developed morphology, random intervals with incorrect spelling and punctuation errors. As there are difficulties with the text processing in these areas, a hidden message based on these vulnerabilities will be very difficult or even impossible to detect. One also shouldn't forget about the ontological approach, the essence of which is to present sentences in the form of Meaning-Text, so the tree of dependencies with semantic relations between the participants of a situation is formed [4, c. 51-52]. Thus, information is concealed on a semantic level, which means that in order to identify the fact of information concealing it is necessary to conduct a comprehensive linguistic analysis of the text in order to detect inconsistencies with general rules and norms, which is difficult when there are problems with the implementation of the textual automatic analysis.

The problem of automatic text processing for steganalysis tasks can be solved by creation of appropriate computer software aimed at counteracting such methods of steganography, as random intervals methods, syntactic and semantic methods. Morphological, syntactic and discourse analyzers, which in the complex can detect traces of modification of the text based on the ontological approach, can counteract these methods, especially concerning discourse analysis. However, appropriate obstacles arise in the way of their creation.

For example, Olha Babina considers the morphological analysis as the basic component, which is the automatic text processing [5, c. 38]. The author notes the difficulties of automatic analysis, which may be caused by the features of a particular language, and features of the automatic analysis algorithm and its implementation [5, c.

38]. The author proposes to solve these difficulties by developing a corpus method of morphological analysis that takes into account the peculiarities of many languages due to the use of a unique algorithm which has the properties of both the dictionary and the non-dictionary systems.

A discursive analysis requires data obtained in the process of parsing, which is impossible without the data obtained in the morphological analysis. So as it is noted in [6] there are problems caused by the syntactic complexity and the process of developing a coherent narration [6, p. 2], which can be solved by effective parsing, with the subsequent allocation of the text meaning.

As it is noted in [7], computer steganography means can be based on linguistic generation methods (NICETEXT, TEXTO) and linguistic modification methods (T-LexSystem) [7, p. 68]. Each category is based on the corresponding vulnerabilities of the text data processing software. Thus, programs of automatic generation create text without meaning. The meaningfulness of text can be recognized only by conducting automated discourse analysis with computer tools. As for linguistic modification, automated morphological analysis and parsing will reveal traces of changes made in the text of natural language.

To solve problems related to the text information analysis automation, software product development is successfully carried out. Thus, there are the most effective systems for the text analysis, which include the syntactic analyzers as AOT, the Russian Context Optimizer program, which is designed to solve the problems of the foreign languages analysis automation, Link Grammar Parser [8], or automated morphological analyzers such as Mystem, LingSoft, or AOT [8]. The mentioned programs solve the main problems of the text data processing automation, but the usual analysis of the text, without taking into account the peculiarities of the stegosystems' construction, cannot solve the problem of data processing automation in the textual steganalysis systems. This is due to the fact that automated text analysis systems, such as mentioned corpus method, are not aimed to detect traces of text modification by steganographic means, based on vulnerabilities caused by the text processing automation problems. So it can be argued that the way of solving the problem of data processing automation for textual steganography is the textual steganalysis software products creation, which apply a comprehensive approach to the textual analysis and it will solve the problem of the

wrong spelling of the words in the text by spelling checking, the problem of different options of writing the same word in the text by morphological analysis, developed morphology of some languages by applying a corpus method of morphological analysis, dependence on the context and the presence of homonyms by the discourse analysis. Also, the widespread steganographic methods, such as methods of random intervals, syntactic or semantic methods should be taken into account. The peculiarities of hermeneutics and semiotics should be taken into account in a morphological, syntactic or discursive analysis, since this will allow to provide a deeper analysis of the text, revealing its basic meaning by using linguistic means, and also will be able to solve the main problems arising in the automation of data processing precisely in light of the textual staganalysis needs.

Conclusions. So, there is a series of difficulties associated with the interpretation of the natural language by the computer program in the process of automation the textual data processing. These difficulties generate vulnerabilities on which most methods of textual steganography are based. Problems related to the text processing automation could be solved during the development of morphological, syntactic or discourse analyzers, most of which are effective in the textual analysis. The software complex, which inherits all the properties of the text analyzers and takes into account the peculiarities of hermeneutics and semiotics for expanding the possibilities of discourse analysis as well as the peculiarities of attacks on the textual stegosystem could solve the problems of data processing automation for the tasks of steganalysis.

References:

1. Nechta, Ivan An Effective Steganalysis Method Based on Data Compression / Ivan Nechta // Siberian State University of Telecommunications and Informatics Newsletter. – 2010. – #1. – pp. 50-55.
2. V. Beloziorov Preparation the Initial Data for Documents' Clusterization / Beloziorov V. // The theses collection of the young scientists' congress, Issue 1. – St. Petersburg National Research University of Information Technologies, Mechanics and Optics. – 2013. – pp. 4-5.
3. Kukharska, Nataliia The textual steganography methods' analysis / Nataliia Kukharska, O. Kryzhanovska [Electronic resource] – Available: https://sci.ldubgd.edu.ua/handle/123456789/753
4. Babina, Olha Linguistic Steganography: Modern Approaches. Part 2 / Olha Babina// The South Ural State University Newsletter. Series: Linguistics. – 2015. – #4. – pp. 49-55.
5. Babina, Olha and Diumin, Nikita The Corpus Method of Automatic Morphological Analysis of Inflectional Languages / Olha Babina, Nikita Diumin// The South Ural State University Newsletter. Series: Linguistics. – 2012. – #25. – pp. 38-44.
6. Michaela Regneri, Diane King Automated Discourse Analysis of Narrations by Adolescents with Autistic Spectrum Disorder In.: 7th Workshop on Cognitive Aspects of Computational Language Learning (CogACLL 2016), Berlin, 2016. pp. 1-9.
7. Chen, Zhili; Huang, Liusheng; Yang, Wei Detection of substitution-based linguistic steganography by relative frequency analysis / Zhili Chen, Liusheng Huang, Wei Yang // Digital investigation. – #8 (1). – July 2011. – pp. 68-77.
8. The linguistic analysis and text processing programs [Electronic resource] – Available: http://www.asknet.ru/analytics/programms.htm

SCIENTIFIC EDITION

MATERIALS OF INTERNATIONAL SCIENTIFIC
AND PRACTICIAL CONFERENCE

_____

«SCIENTIFIC DEVELOPMENT AND ACHIEVEMENTS»

(1 December 2017, St Andrews, Scotland, UK)
_____

In Ukrainian, Russian and English

Materials published in author's edition
Organizing Committee did not always agree with the position of authors
For the accuracy of the material authors are responsible