APPLICATION of INFORMATION and COMMUNICATION TECHNOLOGIES - AICT2015

CONFERENCE PROCEEDINGS

14-16 October 2015, Rostov-on-Don, Russia

www.aict.info/2015

APPLICATION of INFORMATION and COMMUNICATION TECHNOLOGIES - AICT2015

CONFERENCE PROCEEDINGS

2015 9th International Conference on Application of Information and Communication Technologies (AICT)

IEEE Catalog Number: CFP1556H-PRT
ISBN: 978-1-4673-6855-1

Additional copies of this publication are available from
Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
+1 845 758 0400
+1 845 758 2633 (FAX)
email: curran@proceedings.com

IEEE Catalog Number CFP1556H-ART
ISBN: 978-1-4673-6855-1

CONFERENCE COMMITTEES:

CONFERENCE CHAIRMEN

Prof. Marina Borovskaya, *Rector of Southern Federal University, Russia*
Prof. Ahmet Sanich, *Rector of Qafqaz University, Azerbaijan*

CONFERENCE VICE CHAIRS

Prof. Inna Shevchenko, *Vice-Rector for Research and Innovation, Southern Federal University, Russia*
Prof. Niftali Gocayev, *Qafqaz University, Azerbaijan*

GENERAL CHAIRS

Assoc. Prof. Abzetdin Adamov, *Qafqaz University, Azerbaijan*
Prof. Vladimir Kureychik, *Southern Federal University, Russia*

TECHNICAL PROGRAM CO-CHAIRS

Prof. H.Levent Akin, *Bogazici University, Turkey*
Associate Prof. Vincent Guyot, *ESIEA/LIP6, France*
Prof. Asoke Talukder, *IIIT, Bangalore, India*
Prof. Cevdet Meriç, *Fatih University, Turkey*
Prof. Serdar Korukoğlu, *Ege University, Turkey*
Prof. In Jeong Lee, *ITIRC Vice-president, Hoseo University, Korea*
Dr. Tae Woon Kang, *National Research Foundation, South Korea*
Assistant Prof. Kshetrimayum Rakhesh Singh, *IIT-Guwahati, Guwahati, Assam, India*
Dr. Nazim Agoulmine, *University of Evry Val d'Essonne, France*
Dr. Nargiza Usmanova, *IEEE Comsoc Chapter chair, TUIT, Uzbekistan*
Prof. Alok Kumar Das, *Guru Nanak Institute of Technology, India*
Professor Bakhyt Matkarimov, *Nazarbayev University Research and Innovation System (NURIS), Kazakhstan*
Dr. Gennady Smorodin, *EMC Academic Alliance Russia & CIS, Russia*

PANEL CHAIR

Assoc. Prof. Abzetdin Adamov, *Qafqaz University, Azerbaijan*

KEYNOTES, TUTORIALS AND PUBLICATION CHAIR

Assoc. Prof. Abzetdin Adamov, *Qafqaz University, Azerbaijan*

LOCAL ORGANIZING COMMITTEE

Prof. Vladimir Kureichik, *Head of Computer Aided Design Department, Southern Federal University*
Prof. Yuriy Rogozov, *Head of Department of System Analysis and Telecommunication, Southern Federal University*
Ph.D. Yuriy Kupriyanov, *University Alliance SAP NIS*
Assoc.Prof. Andrey Lezhebokov, *Computer Aided Design Department, Southern Federal University*
Assoc.Prof. Elmar Kuliev, *Computer Aided Design Department, Southern Federal University*
Assoc.Prof. Dmitriy Zaporozhets, *Computer Aided Design Department, Southern Federal University*
Ph.D. student Lada Rodzina, *Computer Aided Design Department, Southern Federal University*
Vladimir Kureichik Jr, student, *Computer Aided Design Department, Southern Federal University*
Anton Orlov, student, *Computer Aided Design Department, Southern Federal University*

ADVISORY AND STEERING COMMITTEE

Niftali Gocayev, *Vice Rector for Research and Science, Qafqaz University, Azerbaijan*
Victor Kureichik, *Taganrog Institute of Technology, Southern Federal University, Russia*
Mehmet Ulema, *Manhattan College, USA*
Adnan Yazici, *Middle East Technical University, Turkey*
Raouf Boutaba, *University of Waterloo, Canada*
Guy Omidyar, *IFIPTC6 WG6.8 Past Chair, USA*
Insepov Zinetulla, *Nazarbayev University Research and Innovation System, Kazakhstan*

TECHNICAL PROGRAM COMMITTEE MEMBERS

Cem Ersoy, *Bogazici University, IEEE ComSoc Chapter Chair, Turkey*
Abolfazl Mehbodniya, *INRS, Canada*
Asoke K Talukder, *IIIT Bangalore, India*
Babek Abbasov, *Qafqaz University, Azerbaijan*
Etibar Seyidzade, *Qafqaz University, Azerbaijan*
Canchi Radhakrishna, *KYOCERA, USA*
Djamel Sadok, *Universidade Federal de Pernambuco, Brazil*
Gholam Ali Rezai rad, *University of Science and Technology, Iran*
Guy Omidyar, *Consultant, USA*
Agasi Melikov, *National Aviation Academy, Azerbaijan*
Nadir Alishov, *Ukrain National Akademy of Science, Ukrain Guy Pujolle, Université Paris 6, France*
Guy Pujolle, *Université Paris 6, France*
Alekber Aliyev, *Baku State University, Azerbaijan*
Ramin Mahmudzade, *Baku State University, Azerbaijan*
Petr Sosnin, *Ulyanovsk State Technical University, Russia*
Khaldoun Al Agha, *LRI, University of Paris XI, France*
Khaled Ben Letaief, *Hong Kong University, HK*
Halil Ismailov, *Qafqaz University, Azerbaijan*
Ali Sahintas, *Qafqaz University, Azerbaijan*
Marie-Jose Montpetit, *Motorola, USA*
Mehmet Ulema, *Manhanttan College, USA*
Masoud Mashhouri, *Carleton University, Canada*
Tugrul YANIK, *Fatih University, Turkey*
Fatih CAMCI, *Fatih University, Turkey*
Nazar Elfadil, *SQU, Oman*
Ruslan Kamaev, *KSUCTA, Kyrgyzstan*
Abdul Halim Zaimİ, *Istanbul University, Turkey*
Kamil Saraç, *University of Texas at Dallas, USA*
Hasan Bulut, *Ege University, Turkey*
Muhammet Cinsdikici, *Ege University, Turkey*
Chandana P. Withana, *Charles Sturt University, Australia*
Ahmet Burak Can, *Hacettepe University, Turkey*
Ramiz Alguliyev, *Institute of Information Technologies of ANAS, Azerbaijan*
Galina Mehdiyeva, *Baku State University, Azerbaijan*
Mehemmed Mehdiyev, *Baku State University, Azerbaijan*
Zafar Cafarov, *Azerbaijan Technical University, Azerbaijan*
Rustem Biyashev, *Institute of Information and Computation Technologies, Kazakhstan*
Gulnara Zhangisina, *Satpaev Kazakh Technical University, Kazakhstan*
Arnur Tokhtabayev, *Gumilyov Eurasian National University, Kazakhstan*
Atakan Huseyin Varol, *Nazarbayev University, Kazakhstan*
Zhandos Yessenbayev, *NURIS, Kazakhstan*
Timur Umarov, *KBTU, Kazakhstan*
Sain Saginbekov, *Suleyman Demirel University, Kazakhstan*
Alexandr Gavrilov, *IBM Science & Technology Center, Russia*
Valeriy Tarasov, *Bauman Moscow State Technical University, Russia*
Yuriy Kravchenko, *Southern Federal University, Russia*
Alexandr Sviridov, *Southern Federal University, Russia*
Sergey Kucherov, *Southern Federal University, Russia*
Daria Zaruba, *Southern Federal University, Russia*
Ali Pourmohammad, *Amirkabir University of Technology, Iran*
Hasan Ogul, *Bashkent University, Turkey*
Anar Rustemov, *Qafqaz University, Azerbaijan*
Fatma Büyüksaraçoğlu, *Trakya University, Turkey*
Mohammad Haghighat, *University of Miami, US*

TABLE OF CONTENTS

SESSION 1 – BIG DATA MANAGEMENT AND APPLICATION

SESSION 2 – DATA MINING AND DATA ENGINEERING

SESSION 3. CYBER SECURITY ISSUES

SESSION 4. LATEST TRENDS IN ICT APPLICATION

SESSION 5. COMMUNICATION, NETWORK AND HARDWARE

SESSION 6. ICT IN EDUCATION, HUMANITIES, SOCIAL SCIENCES AND RESEARCH

2015 9th International Conference on Application of Information and Communication Technologies (AICT)

Academic Papers Evaluation Software

Viacheslav Lanin
Business Informatics Department
National Research University Higher School of Economics
Perm, Russian Federation
vlanin@hse.ru

Yuliya Shuchalova
Computer Science Department
Perm State National Research University
Perm, Russian Federation
shuchalova@gmail.com

Svetlana Strinyuk
English Department
National Research University Higher School of Economics
Perm, Russian Federation
sstrinyuk@hse.ru

Abstract

Mastering academic English is crucial for both academicians and students functioning in English speaking environment. However, getting the command of the language at the advanced level might be a long and a painful process. Creating software providing analysis of academic English can play an important role in addressing the issues of linguistic analysis of academic discourse as well as in teaching English for academic purposes. The paper describes the results of development and implementation of the special software to analyse academic texts. Categories are suggested for identifying most important features of academic discourse.

***Index Terms*— academic paper style, GATE, style markers, natural language processing**

I. Introduction

Writing style of an academic paper plays a pivotal role in presenting research results. To be published scholars have to meet strict requirements of scientific journals. Researchers who are not native speakers of English struggle through manuals and guidelines for academic writing, however materials are often rejected due to low quality of writing. Special courses in EAP or picking up language from academic papers are not always a good remedy due to natural limits of time and effort spent to master academic writing. Educational environment for those who would like to become fluent in academic writing could solve this problem.

This work presents an approach to style analysis of academic articles, which is based on extraction of academic style markers and following statistical processing of data. The authors assume that evaluating the "quality of academic discourse" of the text in terms of style can be automated by using software to tag style markers in that text. Creating a repository of patterns is at the heart of this approach but it needs most serious attention. The patterns are needed to extract the markers mentioned above. Evaluation of statistical bounds of markers occurrence requires using methods and tools of corpus linguistics. The implementation of the software product is based on a wide range of capabilities for text corpus processing provided by GATE Developer [1] software, which is used to annotate the text.

Research results are to be used in academic writing analysis to identify most significant features of academic discourse; it may be helpful for scholars who are not native speakers of English, finally, it will allow to evaluate the quality of papers written by students against a number of standardized formal criteria. It may also be published as a publicly-available service for comparing user-provided text with text corpora, provided by the system.

II. Problem Statements

Working on this project implies development of a software solution that will provide the following features:

- annotating text with style markers, described in a repository;
- visualization of extracted markers, including search/navigation options;
- corpus storage and statistical information about style marker occurrence in the particular corpus;
- comparison of deviation between text and corpus measures;
- providing a user with recommendations on style improvement;
- capability to dynamically enhance the set of extracted marker types;
- providing a visual language for describing lexical-syntactic patterns for analyzed marker types.

III. Related Works

Currently there are several commercial and research solutions for stylistic control of both artificial and natural languages. Solutions for source code revision are implemented in programming tools [2]. StyleCop analyzes C# source code to enforce a set of style and consistency rules. It can be run from inside of Visual Studio or integrated into many other

development tools. StyleCop provides value by enforcing a common set of style rules for C# code. Developers can either implement their own rules or use a set of rules provided by StyleCop initially.

The most notable tools for natural languages are systems for professional editors.

Usually in natural language processing features, also known as style markers, are used to define document's style. [3] describes following possible feature types:

- token-level features (e.g., word length)
- syntax-based features (e.g., part-of-speech tags);
- features based on vocabulary richness (e.g., type-token ratio);
- common word frequencies.

These features are used in many machine learning classification methods (listed in [4]), in which each document is associated with a corresponding feature vector.

One approach of academic style analysis has been presented in [5]. All evaluated features are divided into sets that represent formal voice, readability and vocabulary of the text, which are then used to create Self-Organizing Maps. This method is used not only for academic style and readability analysis, but also for authorship attribution.

Another method was used to define characteristics of article titles in computer science. Features, such as title length, punctuation and preposition usage, word frequency, were investigated within a corpus of articles for science journals.

The approach given in this paper is also based on corpus linguistics and involves comparison between academic papers, defined as high-quality, and papers that implement some of the recommendations from various study materials, but still do not apply to the standards and their readability is considered low. Features that are used for the analysis are selected from academic style guides and other methodical literature. The significance of each feature will be evaluated by comparing feature vectors of reference papers from reviewed sources and high- and low-quality student papers.

Frequency of feature occurrence can be measured using lexical-syntactic patterns – declarative structures, which represents its lexical, basic syntactic and also sometimes semantic qualities [6]. These patterns are often used to extract key phrases, build ontologies from text corpora, perform automatic annotation, etc. Lexical-syntactic patterns are usually constructed by philologists or linguists for different tasks and languages.

A common method of describing lexical-syntactic patterns is using regular expressions, which may contain [7]:

- lexical features (token name, category, root, etc.);
- logical operators (OR, AND, NOT);
- special symbols ("wildcards");
- value assignment for variables;
- grouping operators;
- repetition and range;
- optional constituents.

This way of representing patterns can be easily implemented, as a plenty of programming languages support working with regular expressions. Other possible solutions include special structures (e.g., JAPE patterns [8]) and specific languages (e.g. two independent languages named LSPL for processing Russian texts [9, 10]).

As it was stated above, the described approach to academic style analysis implies working with text corpora. There are different kinds of tools to analyze and process corpora, many of which are presented in [11]. Those tools can be roughly divided into following groups:

- concordancers, which are mainly used by linguists to analyze contexts of phrase occurrence and solve several accompanying problems, such as keyword extraction;
- tools for manual annotation, also used to collect statistics and search information based on annotations;
- tools for vocabulary analysis, for example, based on word lists;
- libraries and special software for natural language processing, used by programmers and computer linguists;
- development environments, which can be used to construct processing resources from available or user-developed plugins.

The last group is the most efficient in the context of current task, which implies automatic annotation and statistics gathering based on customized features. The software chosen for implementation of the system is GATE (General Architecture for Text Engineering) [12]. GATE provides a development environment with many basic processing resources (e.g. tokenizers, sentence splitters, morphological taggers) as well as an object library that can be used to write plugins specifically for the task.

IV. Academic style markers

Writing research papers in English requires using several standard linguistic means that are necessary and/or appropriate in academic writing. At the same time usage of those means should be balanced to maintain quality and readability of the text. Recommendations on frequency of style markers occurrence are rarely given in style guides, so the analysis of revised papers may provide results useful for non-native English speakers.

The theoretical foundation of the system described in this paper consists of a list of style markers that were selected from reference and study materials, Internet resources about academic writing as well as scientific papers on this topic. All markers from this list can be divided into three main groups:

- lexical markers;
- grammar markers;
- syntactic markers.

Lexical markers include three types of features:

- specific words and terminology (high frequency of terminology; usage of abstract semantic verbs,

desemantisized verbs, intensifying adverbs; low frequency of personal pronouns *you, he, she*, etc.);
- words corresponding to specific word-formation constructions (nouns with *-or* suffix, commonly used in terminology; abstract nouns derived by suffixes *-ment*, *-ness*, *-tion*, etc.);
- words of specific part of speech (high frequency of nouns, low frequency of pronouns).

Two types of features that fall into *grammar markers* category are:
- wide usage of verbs in Passive Voice;
- presumable prevalence of verbs in Present Tense.

Syntactic markers can also be classified into two types:
- features described by syntactic structures (simple, complex or compound sentence structure; prepositive and postpositive attributes by most of the nouns; possible prevalence of prepositive attributes in technical texts);
- specific conjunctions, linking expressions, etc. (subordinating and correlative conjunctions; archaisms *thereby, therewith, hereby*; prepositional phrases; means of logical expressions).

Most of these features can be automatically annotated using lexical-syntactic patterns, although absolute accuracy cannot be guaranteed, which is why expert control and means of manual annotation correction is highly desirable for the system implementation. Flexibility of the system components is also important for development and further testing and debugging due to specificity of academic style feature tagging and natural language processing in general.

Currently our system annotates text based on all of the described style markers with the exception of terminology and sentence structure. Although some components are still being tested, recent resulting annotation sets provide enough information to analyze academic writing and deepen the studies about some of the features.

V. Example

Some results of the automatic annotation of the system can be illustrated with an example. The system is being tested on a corpus of open source academic papers published in various reviewed journals (e.g., Springer). All chosen papers come under "computer science" category, classified by the publishers. In this paper we are going to analyze style markers found in a part of one of those papers [13], presented in Fig. 1.

> Attempting to measure the degree to which these publications reflect the mission priorities and funding decisions of OER is beyond the scope of this analysis.
> For the purposes of this study, an "OER-supported article" is defined as any article that has received financial, logistical, or other support from OER to gather data for or to perform all or part of the analysis described in the article; any article that utilizes specimens, data, imagery, etc. collected on an OER-supported expedition; or any article authored or coauthored by an employee of OER.

Fig. 1. Sample of text analyzed by the system

Before style marker tagging, the text undergoes tokenization, sentence splitting and POS-tagging, performed using OpenNLP [14] plugins and GATE Morphological Analyser [15]. Those plugins are included in GATE Developer 8.0 package and in theory can be replaced with plugins that annotate text with "token" and "sentence" annotation types and add "string", "category" and "root" features to "token"-type annotations.

This sample of text contains 9 types of different style marker annotation types: abstract semantic verbs, abstract suffix nouns, desemantisized verbs, nouns, Passive Voice, Past and Present verb tenses, postpositive and prepositive attributes. Detailed description is provided below:

- From the given lists of abstract semantic verbs and desemantisized verbs only the word *be* occurs in this text ("funding decisions of OER *is* beyond the scope"). It should be noted that *is* as part of other verb phrases does not count as an independent verb (e.g., in "an "OER-supported article" is defined as...").
- There are several nouns with abstract suffixes, such as *-tion/-ion*, *-y*: *publications, mission, priorities, decisions, imagery, expedition*. The plugin responsible for this type of annotations provides an easy way to extend the list of searched suffixes and word categories, if needed.
- "Noun" annotations depend fully on the results of POS-tagging by OpenNLP plugins. In this sample all of the nouns were identified correctly: *degree, publications, mission, priorities, funding, decisions, OER, scope, analysis, purposes, study, acrticle, support, data, part, specimens, imagery, expedition, employee.*
- Passive voice in the role of predicate occurs once in this text: an "OER-supported article" *is defined*.
- Tagging of verbs of different tenses actually proves the imperfection of automatic annotation. While verbs in present tenses were extracted correctly ("publications *reflect"*, "funding decisions of OER *is* beyond the scope", "an 'OER-supported article' *is defined"*, "article that *has received",* "any article that *utilizes*"), one word is identified by a POS-tagger as a verb in past tense, when in fact it has a role of a participle ("any article that utilizes specimens, data, imagery, etc. *collected* on an OER-supported expedition").
- Prepositive attributes found in this text are as follows: *mission priorities, funding decisions, OER-supported article, other support, OER-supported expedition*. Postpositive attributes: *decisions of OER, support from OER, OER to gather, analysis described, article authored, employee of OER*. Not all of them are found correctly, for example, *to gather* is definitely not an attribute of *OER* in this case, even though such

attribute constructions are possible. On the other hand, *scope of this analysis*, which is a noun with postpositive attribute, was not identified. Also there are situations where a noun is surrounded by both prepositive and postpositive attributes, as in "funding decisions of OER", where *funding decisions* and *decisions of OER* actually form one phrase.

This example illustrates some important features of academic texts, such as relatively large amount of nouns (27 out of 91 word-type tokens – about 30%), rather frequent usage of passive voice (here – once in two sentences), prepositive and postpositive attributes by most of the nouns.

VI. Effectiveness evaluation

The system is currently implemented as a research prototype and undergoes expert testing and further improvement. On each iteration of testing cycle (automated and expert annotations' comparison – fault analysis – re-development) the accuracy of annotation is improved and some rules of annotating are changed to correspond to linguistic study objectives.

To illustrate effectiveness formally the values of precision and recall will be used. These terms are taken from information retrieval [16]. In case of corpus annotation (within one type of style marker annotations) *precision* can be defined as ratio of the amount of correct automatically applied annotations to all automatically applied annotations, while *recall* can be defined as ratio of correct automatically applied annotations to all annotations, manually applied by experts.

At the present time many features that are based on word lists (such as abstract semantic verbs, abstract suffix nouns, intensifying adverbs, etc.) show highly accurate results, where faults happen mainly because of confusion in part-of-speech tagging (e.g., mistaking participles for verbs). Such faults so far have only occurred in some texts and have negative impact only in terms of precision, but not recall.

Prepositive and postpositive attributes are the most problematic features to extract, because they require constructing lexical-syntactic patterns that are not always precise and adequate and demand constant reworking and additions. Attributes that consist of several sequential constructions that can be identified as different attributes raise special concern and need to be studied more closely. Also some attributed include list of coordinate objects and conjunctions, which requires new lexical-syntactic patterns. Currently both precision and recall of these types of attributes exceed 80%, but extraction of these features, as well as some others, is going to be improved in the near future.

VII. Conclusion

To this date the main architectural solutions for the implementation of the project have been selected and basic pattern for further extraction of analyzed markers are constructed. At a next step of our research it is planning to use multidimensional document ontology (MDO) for parsing arbitrary papers in any formats. In addition at a later stage a visual language for pattern description is going to be developed to implement dynamical enhancement of style marker repository and the analysis service is going to be published on open access.

Acknowledgment

The reported study was supported by RFBR, research project № 14-07-31273 mol_a and 14-07-31330 mol_a.

References

[1] Cunningham H., Maynard D., Bontcheva K. Text Processing with GATE. – Gateway Press CA, 2011.

[2] StyleCop https://stylecop.codeplex.com/documentation

[3] Luyckx K., Daelemans, W. Shallow text analysis and machine learning for authorship attribution // Computational Linguistics in the Netherlands 2004: selected papers from the Fifteenth CLIN Meeting / van der Wouden T. [Ed.], e.a., Utrecht, LOT, 2005, – C. 149-160.

[4] Ермакова Л. М. Методы автоматической классификации текстов по функциональным стилям / М. А. Абашев, Р. В. Никитин, Р. И. Ушаков // Вестник Пермского университета. Математика. Механика. Информатика. Выпуск 4(27). – Пермь, 2014.

[5] Scholz T., Conrad S. Style Analysis of Academic Writing // Natural Language Processing and Information Systems: 16th International Conference on Applications of Natural Language to Information Systems, Proceedings. NLDB 2011, Alicante, Spain, June 28-30, 2011. – C. 246-249.

[6] Большакова Е.И. Лексико-синтаксические шаблоны в задачах автоматической обработки текстов / Н.В. Баева, Е.А. Бордаченкова, Н.Э. Васильева, С.С. Морозов // Компьютерная лингвистика и интеллектуальные технологии: Труды Международной конференции Диалог ‘2007 / Под ред. Л.Л. Иомдина, Н.И. Лауфер, А.С. Нариньяни, В.П. Селегея – М.: Изд-во РГГУ, 2007. – С. 70-75.

[7] Jacobs P.S., Krupka G.R., Rau L.F. Lexico-semantic pattern matching as a companion to parsing in text understanding [Электронный ресурс] // URL: http://pdf.aminer.org/000/511/641/lexico_semantic_pattern_matching_as_a_companion_to_parsing_in.pdf (дата обращения: 10.01.2015).

[8] GATE JAPE Grammar Tutorial Version 1.0 [Электронный ресурс] // URL: https://gate.ac.uk/sale/thakker-jape-tutorial/GATE%20JAPE%20manual.pdf (дата обращения: 10.01.2015).

[9] Большакова Е.И., Васильева Н.Э., Морозов С.С. Лексико-синтаксические шаблоны для автоматического анализа научно-технических текстов // Десятая Национальная конференция по искусственному интеллекту с международным участием КИИ-2006. Труды конференции в 3-х томах. Т. 2. – М.: Физматлит, 2006. – С.506-524.

[10] Рабчевский Е.А. Автоматическое построение онтологий на основе лексико-синтаксических шаблонов для информационного поиска. // Труды 11-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL’2009. – Петрозаводск, 2009. – С. 69-77.

[11] Software, Tools, Lists, Resources [Электронный ресурс] // URL: http://www.uow.edu.au/~dlee/software.htm (дата обращения: 05.01.2015).

[12] Linguistica | SourceForge.net [Электронный ресурс] // URL: http://sourceforge.net/projects/lxa/ (дата обращения: 08.01.2015).

[13] Belter, C. W. A bibliometric analysis of NOAA's office of ocean and exploration research // Scientometrics, 95(2). с. 629–644.

[14] GATE.ac.uk OpenNLP Plugin [Электронный ресурс] // URL: https://gate.ac.uk/sale/tao/splitch23.html#sec:misc-creole:opennlp (дата посещения: 05.06.2015)

[15] GATE.ac.uk GATE Morphological Analyzer [Электронный ресурс] // URL: https://gate.ac.uk/releases/gate-6.0-build3764-ALL/doc/tao/splitch17.html#sec:parsers:morpher (дата посещения: 05.06.2015)

[16] Агеев М.С., Кураленок И.Е. Официальные метрики РОМИП'2004. // Российский семинар по Оценке Методов Информационного Поиска (РОМИП 2004) Пущино, 2004.

[17] Strinyuk S. A., Lanin V. Analysis System of Scientific Publications Based on the Ontology Approach, in: Supplementary Proceedings of the 3rd International Conference on Analysis of Images, Social Networks and Texts (AIST'2014). Ekaterinburg : CEUR Workshop Proceeding, 2014. P. 81-85.

[18] Lanin V. Document analysis based on multidimensional ontology of electronic documents, in: Advances in Information Science and Applications. Volumes I & II. Proceedings of the 18th International Conference on Computers (part of CSCC '14) / Отв. ред.: E. B. Zamyatina. Vol. 1-2. Santorini Island : CSCC, 2014. P. 524-527.