

Method for Generating Subject Area Associative Portraits: Different Examples

I. Galina¹, M. Charnine¹, N. Somin¹, V. Nikolaev¹, Yu. Morozova¹, O. Zolotarev²

¹Institute of Informatics Problems, Federal Research Center,
Russian Academy of Sciences, Moscow, Russia

²Russian New University, Moscow, Russia

Abstract - For the last several years a group of scientists has been working on creation of a knowledge extraction system based on the automated generation of subject area associative portraits (SAAP) and on the construction of semantic context spaces (SCS) [1-11]. The ideology of the SAAP is based on the distributional hypothesis, stating that semantically similar (or related) lexemes have similar context and, conversely, similar context has similar lexemes. In the applied model we use an extended hypothesis, that includes not only the study of similarities and differences in the contexts of individual lexemes, but also arbitrary multiple lexemes fragments (significant phrases - SP) too.

Keywords: Subject Area, Associative Portraits, keywords extracting, significant phrases, thesaurus, Big Data

1 Introduction

As a theoretical model of the SAAP in this draft adopted semantic context space (SCS). Formally, a SCS is defined as a graph $G=(V,E)$ with nodes v of V of important terms/phrases from subject area (SA) and the arcs of the graph $(v_i,v_j,Link,w_{ij}) \in E$ describing the relationship/connection between the phrases, where w_{ij} is the weight that expresses the power of association, and the Link-type of connection determined by the type of lexical and syntactic pattern/design, linking phrases. Informative, SCS is a dictionary of significant phrases, the elements of which are connected by associative relations. Any SAAP is a concrete implementation of SCS [1].

The project is implemented on very large data volumes (big data), presented in the Internet environment, which significantly increases the quality of the formed SAAP. As a material research we use freely distributed Russian-speaking NL-texts from the Internet, as well as for some of the subject areas - English language texts. In the future we plan to expand the project to other languages.

To date, we have processed the following subject areas: in Russian - "Remote sounding of agricultural land", "Psychology of business communication"; "Business processes of enterprises"; "Socio-political portrait of Russian regions" (as the training data for this subject are taken a part of SA one of Federal region "Socio-political portrait of

Tatarstan". Information sources for this topic are: official websites of state institutions, political parties; national and regional media; social network Vkontakte, Twitter microblogging, etc.); "Monitoring of public opinion in the socio-political sphere", tapering to a training sample "Protest activity"; and AUV (Autonomous underwater vehicles") in Russian and English languages; "Computer Science" (in English), etc.

On these topics we found and processed from ten thousand to one and a half million documents in Russian and English languages, a total of about 160 GB. For example, on the SA of "Socio-political portrait of Tatarstan" formed the texts digest of about 20 GB, and on the subject of "Protest activity" there was found over 28 GB of test information.

2 Stages of project

This project is created in several stages. There was developed a version of algorithm formation of SAAP on the first stage, based on the material from different subject areas. The key component of a comprehensive methodology was created and would be developed. This is the module for statistical analysis for the algorithm of SAAP formation. There were performed: thematic searches of NL-texts downloaded from the Internet, the formation of large corpora of these texts, divided by subject areas, with highlighting of significant phrases (SP) and the creation of SP dictionaries in an automated mode. All of these items are composed to a united iterative algorithm which represents a method of obtaining SAAP for any subject area. Some stages will be corrected and improved in further work.

In the second phase of the project, a methodology for automatic generation of the SAAP has been further developed in the form of corrected mechanisms for semantic search in the space of NL-texts from the Internet. The algorithms methods were developed due to the training module of the statistical analysis on test samples with subsequent testing on real text flows from different subject areas. This project is based on the KEYWEN encyclopedia concepts [4].

The author's experiments were conducted on above mentioned test samples; also the plan for further experiments for the entire term of the project was developed. The work will be continued.

At present time, the technique allows to automatically extract keywords (single words-tokens, terms and SP) from NL-texts,

to identify associative links between them, to set an hierarchy of terms and SP with the clustering method and to build dictionaries of SP in various subject areas. The output of the algorithm is informative "cloud" of associative relationships for automatically allocated tokens (terms) and SP, from which the associative portraits of the subject areas (SAAP) are formed. Actually we received a "bunch" of associative links defined the value of the SP term and his place in the terminological system of the studied subject area. The associative links allow us to classify the considered term and/or SP to a particular theme

3 The SAAP forming method

The essence of the proposed method of forming SAAP is to iterate extension of the initial dictionary of SP to full SAAP. The method can be described as the following:

1. Choice of key terms that specify the subject area is carried out by the user: a) manually (at the first stage of the training sample, or to explore new topics); b) can be extracted automatically by the system based on the frequency characteristics (weights of the terms); or c) can be automatically taken out of the previous SAAP as initial AP (keywords).

2. Semantic search on a given keywords and an accumulation of relevant base of Internet texts which make up the corps of NL-texts from different SA (big data). During the debugging and improving the results the machine learning techniques are used.

3. Automatic extraction of terms from texts and drafting of frequency dictionaries of SA (statistical methods).

4. Formation of lists of the most important SA concepts on key terms and SP.

5. Splitting texts into segments or suggestions.

6. Construction of the context of each term as a set of containing segments/sentences.

7. Building of context vectors, i.e. a calculation is build for each meaningful term (or SP), and vector is formed from the obtained statistics. These context vectors determine the measure of proximity and associative relationships between keywords.

8. Calculation of the cosine proximity measure of context vectors (the vector components of SP - key words of this SA - is the joint frequency of occurrence of this SP with other SP's in the same context) and the choice of the most strong associative links. To calculate the cosine measures the following formula is used (Fig. 1):

$$\frac{x \cdot y}{|x||y|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

Fig. 1.

where x and y are vectors in a sign space, which themselves are SP from a given semantic context space (SCS), and i - is an index running over the number of signs.

9. Selection of candidates for the key terms (to expand the original frequency dictionary of SP) from the receipt of the strongest associations in key terms conducted as follows:

- searching for a cluster of closely related keywords;
- building an hierarchical structure on the set of the SPs with the hierarchical clustering method;
- calculating the center of the cluster specified keywords (SP) related to the subject area in the vector space;
 - determining the most powerful terms that are candidates for inclusion in the SP from the number of new terms that are closer to the center of the cluster;
 - setting the most distant from the center of the cluster keywords (SP) as candidates for exclusion from the composition of the SP;
 - accepting the final decision on changing the composition of SP by adding or removing terms-candidates.

10. Replenishment of the initial composition of key terms with the strongest candidates that were automatically chosen by the system and go to step 1.

This is a layout algorithm. The statistical module was developed on the basis of this algorithm. It solves the problem of creating SAAP [1, 11]. The individual steps of the methodology, including mechanisms of semantic Internet surfing will be developed and improved in the next phase of the project.

4 The essence of machine learning

The processing of large amounts of texts from the Internet allows to collect the necessary statistical data to form a fairly complete picture of the SA, available in the form of SCS. The opportunity to perform machine learning on a large number of examples gives the system some flexibility and improves outcome.

Machine learning in identifying the key words of SA and associative relationships between them are based on the following scheme:

- 1) creation of training and testing samples for each text (document) with specified class;
- 2) preparation the texts to use them for learning: breaking the texts into elements - words, SP, punctuation, etc.; identification of "noise" in the text and its clipping; for some training samples lemmatization is done (bringing words to a normal form);
- 3) converting each document into a feature vector;
- 4) creating dictionary of SP (classifier) and further training of the classifier on the obtained training set;
- 5) verifying the accuracy of the classifier on the test array of texts.

Here are some examples of terms in the SP dictionaries denoting the various elements of business processes [5, 9].

A part of the SP dictionary on "Business processes in the enterprise" [12-16]

- names of business processes (repair, inspection, replacement of parts, parts fabrication, delivery);
- input (ordering, fault list, blank invoice, commercial invoice...);

- management (schedule, drawing, regulations, invoice...);
- participant/mechanism (master, mechanic, turner, accountant, driver...);
- output (repaired auto, fault list, item, invoice, goods...).

5 Socio-political portrait of Russian regions

The improvement of the formation algorithm of the SAAP is carried out, including machine learning techniques, with the conducting of experiments in building SAAP. As a training topics to address this problem was chosen subject area "Socio-political portrait of Russian regions". Republic of Tatarstan was selected as a training sample. When building the ontology on a given SA we represent the most significant named entities of the following classes:

- geographical names of major settlements in the region;
- persons, their positions in the administration of the region;
- the most important organizations in the region;
- the most important events in socio-political life of the region;
- levels of power in the region;
- the scope of activities in the region.

The development of mineral resources (oil, coal, gas) for Tatarstan is very important, so as certain types of economic activity in the region. So we got several additional classes in ontology: "petrochemicals", "engineering", "agriculture", "education".

Basic classes of entities, their attributes and relationships are represented on the figure below (Fig. 2):

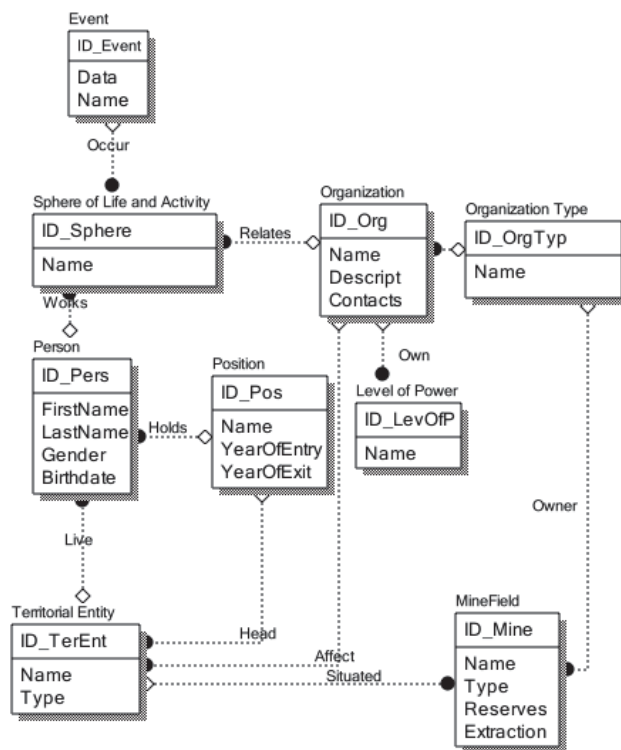


Fig. 2.

The use of ontology in the information system allows us to solve several different problems: 1) building of etalon entities to serve as models for comparison and identification of entities extracted from publications on specific SA originating

from different information sources; 2) creating entities that can be presented to the user as a predefined query object to an integrated data warehouse; 3) constructing relations of reference entities that are represented in the ontology. It allows us to complement and summarize the information contained in the texts. For example, if in the text is found the reference person, it is possible to carry his name to the organization where he works (from ontology), and then generalize to the level of authority and scope of activities.

The low labor cost of automated way of creating ontologies is an advantage. The disadvantages are the lack of completeness of the ontology and sometimes random nature of allocated objects. A fully automated method of producing such ontologies is not possible yet, it still needs some work. While building the initial keyword set, analyzing the material, creating the ontologies for different training samples we had to do some manual editing (several iterations on the results of machine learning).

To improve the completeness and accuracy of the results this work was carried out using the automatic Internet surfing as a result of round the clock operation of the two servers and using crowdsourcing forces of our group, and with the involvement of undergraduate students of MTUCI [6, 10]. Crowdsourcing allowed us to significantly expand thematic corpora NL-texts formed at the initial stage. This work was carried out according to the following scheme:

- the system monitors the relevant information on the Internet of some SA;
- the user specifies a search query in the form of a key terms set (words and phrases);
- the analytical system finds some phrases on the Internet-site with specified key terms and fixes links (URL) to the relevant documents;
- out of found sentences automatically are extracted new SP that allow to expand the original search query and provide greater completeness of the information retrieved.

6 The constructing of search query

Improvement of surfing mechanisms is a necessary condition for the successful extraction of knowledge from natural language texts of very large scale (big data); improvement of these mechanisms is a separate subtask.

The used search algorithm is partially inherited from the electronic system - the encyclopedia of keywords Keywen; this algorithm produces as an intermediate result texts digests of specific subject area and looks like this. The search query is formed on the base of plurality significant phrases (SP). The search of all combinations of key terms is performed (singular terms, the terminology pairs, triples, etc.) so that the length of the query does not exceed a specific limit "n". The average number (several hundreds) of relevant documents must be found as a result. The n-value is determined

empirically. Then search queries are processed with the use of well-known search engines (Google, Yandex and other) and as a result we get a set M1 of text documents. This set is viewed concerning granting of URLs. The extended set of documents M2 is formed on these links. Further, the documents from the set M2 are divided into sentences (or fragments, similar in length to the usual offerings). As a result a database is compiled with entries of the following form:

<text fragment>-<URL pointing to this fragment>

Such a database has a significant amount of volume (up to several TB). For its storage and retrieval, it is not enough traditional means for working with databases. You must use special software and hardware technologies of Big Data processing for the purpose. Statistical estimates are obtained on the basis of the data. We calculate of how many different URLs refer to the same sentence. Thus, each fragment receives a rating equal to the number of its occurrences on independent sites. From the sorted top by rating fragments (by a certain procedure editing) is made a kind of texts digest containing rating information on this SA (without the URL).

7 Socio-political portrait of Tatarstan as an example of search query

Here are some examples of SAAP methods in various subject areas, including the resulting texts digests. Subject areas of "Socio-political portrait of Tatarstan" and "Protest activity" were taken as training samples.

Subject area: "Socio-political portrait of Tatarstan"

To create texts digest - training sample by Socio-political portrait of Tatarstan was used a search query consisting of the two components: 1) territorial formation of Tatarstan and 2) words that are characteristics of texts on social and political life. The work result of surfing semantic system was the digest of the sentences that satisfy the search query with over 40,000 sentences. Lists of officials and organizations of Tatarstan were received. These lists have been manually edited to exclude objects that are not related to the study SA.

Search query

(Tatarstan OR the Republic of Tatarstan OR Tataria OR Agryz OR Aznakaevo OR Almetyevsk OR Arsk OR Bavly OR Bugulma OR Buinsk OR Elabuga OR Zainsk OR Zelenodolsk OR Kazan OR Laishevo OR Leninogorsk OR Mamadysh OR fee OR fee OR Naberezhnye Chelny OR Nizhnekamsk OR Nurlat OR Tetyushi OR Chistopol OR Agryz district OR Aznakaevo district OR Vysokogorsky district OR state district OR Alexeyevsky district OR alkeevsky district OR Almetyevsk district OR Apastovo

district OR Arsky district OR atna district OR Bavlinsky district OR Baltasi district OR Bugulma district OR Buinsk district OR Cherepovets district OR Vysokogorsky district OR erykla OR Elabuga region OR Zainsk district OR Zelenodolsky district OR kaibitzky region OR Kamsko-Ustinsky district OR Kukmor district OR Laishevsky district OR the Leninogorsk district OR Mamadysh district OR Mendelev area OR Menzelinsk district OR its administrative district OR Nizhnekamsk district OR Novosheshminsky district OR Nurlat district OR Pestrechinsky district OR fish-Sloboda district OR Sabinsky district OR Sarmanovo district OR Spassky district OR Tetyushi district OR Tukai district OR tyulyachinsky district OR Aksubaev OR Aktyubinsk OR Alekseevskoe OR Apastovo OR Baltaci Rich OR subs OR Vasilevo OR Jalil OR Kamskie Polyany OR Kamskoe Ustie OR Karabash OR Kuybyshevskiy Zaton OR Kukmor OR Lower Elm OR lower Maktama OR Rybnaya Sloboda OR Tenisheva OR Urussu) AND ("rating of heads of regions" OR "regional head" OR "public opinion" OR "service of state statistics" OR "national monitoring service " OR " governors rating " OR "the head of the region" OR "social wellbeing" OR "standards of living" OR "potential of the region" OR "the image of the Governor" OR "employment" OR "effective Governor" OR "media efficiency" OR "the economy")

A fragment of the texts digest

Sozinov held meetings with students and their parents in the cities of Republic of Tatarstan, Naberezhnye Chelny, Nizhnekamsk, Yelabuga. Due to numerous requests of consumers in a list of cars sold under the program for the disposal, was added long-haul tractor KAMAZ M (5490). There is a railway line Akbash-Agryz in the Elabuga district (railway station Tikhonovo 14 km South-West of the city), the highway Kazan-Ufa, Elabuga-Mozhga.

As a result a set of texts has been found. High rank persons in the administration of Tatarstan Republic were highlighted.

Partial list of persons in Tatarstan

Person	Position
Rustam Minnikhanov	President of the Republic of Tatarstan
Ilsur Metshin Raisovich	the head of Kazan, the mayor, the Chairman of the City Duma, The Counsellor
Mintimer Sharipovich	former President of the Republic of Tatarstan
Andreeva Lyudmila Nikolaevna	Deputy Chairman of the City Duma
Gabdullina Rosalia Mirzaeva	Chairman of the Board of LLC "Kama commercial Bank"
mirgalimov Hafiz	The head of the faction of the Communist party of Tatarstan, member of the state Council of Tatarstan

A partial list of organizations of Tatarstan

State Council (state Council), Administration of Agryz, Administration Aznakayevo, Administration Almetyevsk Administration Arsk... Administration Chistopol, Agryz district Administration, the Administration Aznakaev area, Administration Vysokogorsky district, the administration of the state district Administration... Chistopol district, Yutazinsky district Administration, Ministry of internal Affairs of Russia Kazan, OMVD Russia, Agryz... MOI of Russia Naberezhnye Chelny... OMVD Agryz district... OMVD Yutazinsky district, the Press service of the MOI of Russia Kazan, Russia, United Russia (UR) Kazan, Russia, Fair Russia, Kazan, LDPR Kazan, the Communist party of the Russian Federation Kazan, ...the Ministry For Civil Defence, Emergencies and Elimination of consequences of Natural Disasters (emergency Department) Russia Kazan,... Russia's Investigative Committee Naberezhnye Chelny,...

The subject area of " Protest activity "

Here is an example of a search query based on the selected keywords used to generate a digest of the subject area "Protest activity" using automatic systems semantic surfing.

Search query

(Activists OR Protests OR the Ruling Elite OR intra-elite Split OR a Coup OR pressure On the Government OR Demonstrations OR Euromaidan OR Seizure of Power OR Conflict of Elites OR the Crisis of Legitimate Authority OR the Maidan OR Mass Protests OR Rallies OR Dissatisfaction OR Public Opinion OR Opposition OR Picketing OR Pickets OR Political Protest OR a Protest Movement OR Protest Behavior OR the Radicalization of Protest OR a Change of Government OR Assembly OR Anxiety OR the Level of Organization OR Value Conflicts OR Processions OR Escalation of Conflict)

On the basis of the request was received this training text sample.

A fragment of the digest from the texts of SA "Protest activity"

This Federal law is aimed at ensuring the implementation of the rights of Russian Federation citizens established by the Constitution of the Russian Federation to assemble peacefully, without weapons, hold rallies, meetings, demonstrations, marches and picketing.

As you know, rallies of community activists and opposition under the name of Euromaidan pass in Kiev and other Ukrainian cities, which were triggered by the government's decision on the suspension of European integration November, 21, 2013.

Here is an example of an intermediate frequency dictionary on the basis of which the decision on the inclusion of the word (phrase) in a Dictionary of SP is made:

A fragment of intermediate frequency dictionary

- The MAIDAN (931)***
- ON THE MAIDAN (359)
- VECHE (171)
- OPPOSITION (591)***
- ACTIVISTS (254)***
- FOLK (92)
- POPULAR ASSEMBLY (89)
- IN KIEV (145)
- ON THE MAIDAN NEZALEZHNOСТИ (87)
- RALLY (331)***
- ACTIVISTS (100)***
- MAIDAN ACTIVISTS (52)
- PROTEST (161)
- OFFENSIVE (25)
- PEACE APPROACH (18)
- ATTACK OF THE ACTIVISTS (15)
- IMPLEMENTATION OF REQUIREMENTS (15)
- DEMONSTRATION (93)***
- PROTESTS (111)***
- THE LEADERS OF THE OPPOSITION (35)
- POWER (149)
- MEMBER (41)
- PROCESSION (47)***
- FRACTIONS (31)

Here is an example of the ontology of the subject area.

The upper level fragment of the ontology of SA "Protest activity":

Objects: events (protests), people (activists, politicians), organization (the party, the opposition, Euromaidan), information sources (agencies, web sites), the time, place.

Protests: protests, demonstrations, picketing, campaign of civil disobedience, strikes, meetings, processions.

8 Conclusion

A tool for the automated generation of SAAP is a promising direction for creating systems of knowledge extraction. In addition, this tool is suitable for solving various other tasks.

A comprehensive approach to automatic processing of NL-texts based on a combination of statistical methods, corpus linguistics and distributional semantics to identify significant phrases (SP) and their association to the formation of semantic context space (SCS). It is implemented in the form of iterative methods of formation of the SAAP, an innovative and relevant research on the Russian language.

The obtained results can be used to solve the problems of the automated semantic search on very large data volumes, to extract knowledge from unstructured NL-texts and for automated generation of interactive domain-specific encyclopedias. Especially promising field of application

SAAP approach is the building of ontology, followed by the construction of database schema and knowledge.

The construction of the SAAP and the SCS using associative links between the selected terms (meaningful phrases), as well as the mapping of terms to the categories calculated by the formula cosine measures, has led to the following results:

- developing the complex of methods of text classification using methods of statistics, distributional semantics, corpus linguistics, machine associative links and some others, which allow to classify texts, initially containing the significant terms and phrases (collocations) from the training set;
- creating the software that implements the proposed methods for the selection of relevant texts on the topic of the subject area;
- forming by the volume of the electronic body correlated with a particular subject area NL-texts with automated partitioning of texts in the subject area;
- creating the body of dictionaries of terms of significant phrases (SP) and subject dictionaries, accumulated by processing large volumes of data from the Internet;
- developing the methods for creating visual maps for different subject areas with the use of multidimensional scaling" technology;
- further development of the method of semantic search NL-texts for the purpose of directed data extraction from network texts (big data) of encyclopedic information.

At present the authors' team continues to work on the development of methods of SAAP and on the creation of software tools for its implementation.

9 Acknowledgements

This work is supported by the Russian Foundation for Basic Research, grant #13-07-00272 "The methods for automatic creation of associative portraits of subject domains on the basis of big natural language texts for knowledge extraction systems".

10 References

- [1] Charnine, M. M., V. Somin, I. P. Kuznetsov, I. Morozova, I. V. Galina, E. B. Kozerenko. Statistical mechanisms of formation of associative portraits of the subject areas on the basis of natural language texts of large amounts of systems for knowledge extraction. // Informatics and its application. 2013. Vol. 7. No. 2. Page.92-99.
- [2] M. Charnine, A. Petrov, I. Kuznezov. Association-Based Identification of Internet User Interests. // Proceedings of the 2013 International Conference on Artificial Intelligence (ICAI 2013). V. I. WORLDCOMP'13. July 22-25, 2013. Las Vegas Nevada, USA. CSREA Press. C. 77-81.
- [3] M. Charnine, V. Protasov. Optimal Automated Method for Collaborative Development of University Curricula. // Proceedings of the 2013 International Conference on Artificial Intelligence (ICAI 2013). V. I. WORLDCOMP'13. July 22-25, 2013. Las Vegas Nevada, USA. CSREA Press. C. 96-100.
- [4] M. Charnine, N. Somin and V. Nikolaev. Conceptual Text Generation Based on Key Phrases. // Proceedings of the 2014 International Conference on Artificial Intelligence (ICAI 2014). Vol.I. WORLDCOMP'14. July 21-24, 2014. Las Vegas Nevada, USA. CSREA Press. C. 639-643.
- [5] O. Zolotarev, M. Charnine, A. Matskevich. Conceptual Business Process Structuring by Extracting Knowledge from Natural Language Texts. // Proceedings of the 2014 International Conference on Artificial Intelligence (ICAI 2014). Vol.I. WORLDCOMP'14. July 21-24, 2014. Las Vegas Nevada, USA. CSREA Press. C.82-87.
- [6] V. Protasov, M. Charnine, and E. Melnikov. The Crowdsourcing Linguistic Technology for Experts Assessment. // Proceedings of the 2014 International Conference on Artificial Intelligence (ICAI 2014). Vol.I. WORLDCOMP'14. July 21-24, 2014. Las Vegas Nevada, USA. CSREA Press. C. 656-661.
- [7] T. N. Borisov, A. E. Bronetz, S. V. Klimenko, V. V. Rykov, M.M. Charnine.. Autonomous underwater vehicles: automatic generation of associative portrait subject area. // Proceedings of International scientific conference "Situational centers and IAS for monitoring and security. SCVRT2013-14". Moscow-Protvino. Publishing house of the Institute of physics and technology (ICPT). 2013-2014. P. 38-43.
- [8] Charnine M. M., Somin N. In., Galina I. V., Rodina I. V. Meaningful generate: a method of generating informative texts of relevant phrases. // Proceedings of the XIX international conference CPT'2014. Larnaca, Cyprus. may 12-18, 2014. Publishing house of the Institute of physics and technology (ICPT) (in press).
- [9] O. Zolotarev A., M. Charnine. Methods of knowledge extraction from natural language texts and the construction of models of business processes on the basis of allocation processes, objects, their relationships and characteristics. // Proceedings of the XIX international conference CPT'2014. Larnaca, Cyprus, may 12-18, 2014. Publishing house of the Institute of physics and technology (ICPT) (in press).
- [10] Johnson, O., M. Charnine. Methods of knowledge extraction from natural language texts and the construction of models of business processes on the basis of allocation processes, objects, their relationships and characteristics. // Proceedings of the XIX international conference CPT'2014. Larnaca, Cyprus, may 12-18, 2014. Publishing house of the Institute of physics and technology (ICPT) (in press).
- [11] Protasov, V. I., Charnine M. M., Melnikov, I. E. the Use of technology self-managed crowdsourcing for certification experts. // Proceedings of International scientific conference "Situational centers and IAS for monitoring and security.

SCVRT2013-14". Moscow-Protvino. Publishing house of the Institute of physics and technology (ICPT). 2013-2014. P. 262-266.

[12] Charnine M. M., Rodina I. V. Mechanisms of formation of linguistic and statistical portrait of the subject area monitoring of public opinion for the technology of big data. // Proceedings of the Third International scientific-practical conference "Social computing: fundamentals, technology, Humanities and social effects" (ISC-14)". September 18-19, 2014 RITZ MGGU. Sholokhov. P. 66-70.

[13] O. Zolotarev, O. Kozerenko E. B., M. Charnine. Principles of construction of models of business processes of the subject area based on the processing of natural language texts // Bulletin of ROSNOU. - M.: RosNOU, 2014. No. 4. P. 82-88.

[14] O. Zolotarev, Process approach to the management of projects in the implementation of corporate information systems. // Bulletin of ROSNOU. - M.: RosNOU, 2014. No. 4. P. 89-92.

[15] O. Zolotarev. Methods of extraction processes, objects, relations from natural language texts. // Security problems of the Russian society. - Smolensk: Svitok. 2014.

[16] O. Zolotarev. Control in implementation projects of distributed Enterprise Resource Planning Systems. Bulletin of the Russian New University, Moscow, RosNOU, 2012. No. 4. P. 78-80.

[17] O. Zolotarev. Innovative solutions in the formation of functional structure of the subject domain. Bulletin of the Russian New University, Moscow, RosNOU, 2013. No. 4. P. 82-84.