# Semantic processor for knowledge extraction from texts in Russian and English

Igor P. Kuznetsov, Elena B. Kozerenko, Nikolay V. Somin

Institute for Informatics Problems of the Russian Academy of Sciences, Moscow, Russia

**Abstract** *The paper is dedicated to one approach to the automatic extraction of the knowledge from natural language texts (Russian and English) with forming the Knowledge Base. It is used for the solution of the most complex problems of the linguistic processors and logical analytical systems. For this purpose the means of knowledge representation (the extended semantic networks - ESN) and the tools of their processing (the language of logical programming DECL) have been designed. On this basis the universal syntactical semantic rules and ontologies have been proposed which are composed of the universal linguistic knowledge for knowledge extraction and which have been used for construction of many intellectual systems for different applications.*

**Keywords** Semantics, Natural Language, Linguistic Processor, Knowledge Extraction, Named Entities

## 1 Introduction

The existing Internet largely consists of unstructured documents. Knowledge contained within these documents can be made more accessible for machine processing by means of transformation into the form to be reasoned with. A more prospective approach consists in using the Knowledge Base. It proposes the development of new technology including the extraction of knowledge structures and organization of their processing in the Knowledge Base [1,2,4].

The distinctive features of our technology are as follows:

1) extraction from texts of knowledge structures (not only separate named entities) that represent the links of named entities and their participation in actions and events;

2) for the knowledge extraction the unique semantic-oriented language processors (LP) are designed; the processor provides deep analysis of NL texts and revealing set of entities together with their structures;

3) the processor LP is controlled by the linguistic knowledge, which are declarative structures (in extended semantic networks - ESN) and which provides the quick tuning of LP to subject areas and languages - Russian and English;

4) linguistic knowledge consists of the rules, which provide the high degree of selectivity in the entities extraction and elimination of collisions during their application; rules provide the minimization of noise and losses, that is the high degree of completeness and accuracy;

5) the knowledge structures and means of their processing (intellectual language DEKL) were designed as the united tools, oriented at the tasks of linguistic analysis, semantic search, logical analytical processing and the expert solutions. Using these tools considerably facilitates the development of applied intellectual systems.

The technology of knowledge structure extraction and processing has been used for construction of new classes of analytical systems [3,7,12,13]: “Criminal”, “Analytic”, “AntiTerror”, “Resume” etc. (http://IpiranLogos.com/en/Systems/).

## 2 Tools for intellectual processing

### 2.1. Extended semantic networks

For knowledge representation the formalism was proposed, i.e. the Extended semantic networks (ESN) [2,3,4]. In the ESN constructions the paradigm is used in which the model of the external world is quantized to the objects and the relationships between them. At the same time the integration of objects is allowed when from simple objects it is possible to build more complex ones. The reverse process is the specification. In each object the parts connected by certain relationships can be selected. This is easily expressed in the natural language (NL) and should be presented in knowledge structures. The extended semantic networks have been designed on the base of this paradigm.

Extended semantic networks (ESN) are composed of fragments of following type:

*<Relation name> (<arg1 "," arg2>, ..., <argN> / <code of fragment>)*

where *<arg1 "," arg2>, ..., <argN>* are the argument places, which may be occupied by constant, or the number of variables, which may correspond to the named entities (information objects). The code of a fragment corresponds to a complex object, i.e. arguments with their relationship, which are considered as a whole. The "relationship" is considered in the broad sense. A unary relation (with one argument) is a property. A binary relation connects the two objects, and an N-ary relation connects more objects. For example, an N-ary relation can be an action in which N objects took part (with different roles).

Codes of fragments are needed to represent the level of integration. A code of fragment may be a constant, which must be "unique", i.e. it cannot be a code of another fragment.

A fragment code may be missing in the record. Then the fragment will take the simpler form:

*<Relation name> (<arg1>,<arg2>, ...,<argN>).*

It is not difficult to see that the fragments have the form of named predicates, where the fragment codes are the unique names of predicates. Many fragments are composed of the extended semantic network (ESN). The order of the fragments in the ESN does not matter. Two features should be noted.

The first feature is the use of the so-called intersystem constants. They are written in the form of numbers with a plus sign (N +), when a constant is introduced for the first time, and a minus (N-), when it has already been mentioned. For example, two fragments NAME(IVAN, 1 +) STRONG (1 -) represent that " *a man named Ivan is strong."* In this case, if we again encounter a number 1 +, it would introduce a new (different) constant. For example, the fragments NAME(IVAN, 1 +) STRONG (1 -) NAME (PETER, 1 +) BRAVE (1 -) WEAK(1 -) are presented two people: "*Ivan is strong, and Peter is brave and weak*." Instead of sign 1+ and 1- there can be any integer (N), i.e. 2 +, 2 -, etc. Intersystem constants are needed to refer to objects that are defined by their properties, relations or presented implicitly. If the text has the two objects named *Ivan*, these can be different people and they are represented in ESN by different constants. It is a difficult procedure to choose their different mnemonics.

The second feature is the following. The codes of fragments (usually intersystem constants) can stand in the argument places of other fragments. It is necessary in the cases when some objects are components of others. For example, the fragments NAME(IVAN, 1 +) BUY (1 -, BOOK / 2 +) DECIDE(1 -, 2 -) are represented "*Ivan decided to buy the book"*, where there are two actions. In this one (DECIDE) includes another (BUY). Every named entity (NE) may be the components of actions or other objects. Because every NE is presented as a fragment of ESN with its own code (see 3.3).

Described features (when some code fragments can be in the other argument places) greatly increases the possibilities of language ESN for representation of different types of information, including the semantic components of NL-construction. They are widely used for describing events and actions by forms with verbal nouns, participial and other constructions. It is significant that these features make the possibilities of language ESN reach far beyond the classical language of predicate logic. For example, ESN make possible the representation of various types of paradoxes which are typical for the NL, but impossible in the logic [1]. Constructions of ESN compose the United Knowledge Base which is used for subject and linguistic knowledge representation and which determines the logical analytical decisions and the work of linguistic processor.

## 2.2. Logical programming language for knowledge processing

For processing the knowledge structures, presented by ESN, a special language of logical programming (DECL) has been constructed [2,3]. The DECL language was used as the base for programming the linguistic processor (LP) which transforms the surface (space) structures of texts to deep (semantic) structures where named entities and their relationships are represented.

The language DECL consists of the rules IF ... THEN ..., called productions. Productions are applied to the knowledge base (KB) and have the form:

*<Name productions> (...): IF <LfP> THEN <RgP>;*

where *LfP* is left part of the production and *RgP* is right part of it. Both parts consist of set of ESN fragments, which (in addition to constants) may contain variables. The fragment <*name productions*> (...) is necessary to call the production application.

The left part (LfP) of a production rule sets the conditions for its application. If the conditions are satisfied (i.e. analogical structures are contained in the KB), then the production is considered to be applicable. As a result the variables in LP take the values and are activated in the right part.

The right side of productions (RgP) determines the actions performed for transformation of structures in the KB. If the productions were applicable, then the actions would be initiated. The values of the variables are transferred from the LfP to RgP and are taken into account in actions.

The parts LfP and RgP contain not only fragments, but also special operators to call productions (by name), and the so-called special fragments (or build-in predicates) that define references to external procedures, for example, the interface programs.

The condition of their application consist in comparing the fragments of the LfP with the fragments of KB. If the corresponding structure in the KB is found, the product is considered to be applicable and values of variables from LfP are transmitted to RgP and take into account in the actions. If RgP has a fragment, then it is added to the KB (http://IpiranLogos/en/Tools/).

# 3. Representation of semantic structures

## 3.1 Type of entities and links for extraction

Named Entities (NE) are extracted from the documents in Natural Language by linguistic processor (LP) and represented in the Knowledge Base as the fragments of the extended semantic network. The arguments of fragments are the collections of normalized words, numbers and signs, which reflect essence of NE and indicate its type.

In our systems more than 40 types of NE are extracted from NL-texts [1,7,8]. Their quantity depends on the subject area and the tasks of users. Let us note that in the KB some NE can be constitutional components of others. Connections between NE may be complicated [1,6,14 ]. We consider that actions with their objects and components are the kind of NE, which are connected by special relations (time, space, reason and so on) with other actions. The ESN have been designed for the representation of such information on a homogeneous base which is necessary for deep computer processing of NL texts – Russian and English [1,10].

The set of the entities to be extracted depends on the tasks of a user. At the same time the quality of a linguistic processor is determined by the possibilities for knowledge extraction. The linguistic processors of the systems “Criminal”, “Analytic” and ‘’Semantix” support more than 40 types of semantic entities which can be extracted automatically.

Standard entities (names, dates, addresses, types of weapons and others) are reduced to one standard form. The identification of entities is performed taking into account brief designations (for example, separate surnames, patronymics, initials), anaphoric references (indicative and personal pronouns, for example, this person, it...) definitions and explanations (for example, the *Mayor of Moscow Sobyanin* is identified with the subsequent words *mayor*, *Sobyanin*). An important task is the identification of entities in the entire text, the use for these purposes of indicative pronouns, brief names, anaphoric references.

The Graph presentation of some extracted entities is shown on fig 1.


Fig. 1 Some extracted named entities

## 3.2. Connections between the entities and participation in actions

*Connections and relations* between NE, extracted from the NL-texts, can be very diverse. They depend on entity types. For example, one person can be connected with another by relative and friendly relations, and also by the place of living, area of interests and so on. Actions frequently are connected with the time and the place. There can be reason-consequence and other connections between actions. In such a way the complex structures are created. For their formalization special tools of knowledge representation have been designed.

*Actions* usually are expressed in NL texts by the tensed verb forms, nonfinite verb forms, e.g. verbal nouns, participial and adverbial constructions, gerunds. The actions are also NE, components of which can be another NE. For example, there can be those, who participate in action, or entities, on which the action is directed. Moreover, some actions may be components of others. For many applications the actions are also the significant information which requires formalization. Because the system is oriented at the deep analysis of text constructions, it extracts all actions and events with NE.

## 3.3. Meaningful portrait of a document

A meaningful portrait of a document is the formal representation of entities (NE), their properties and the connections, extracted from the text of the document. Such portraits are the structures of knowledge. As means of formalization in our technologies we use the extended semantic networks. Formalization is achieved automatically by the semantics-oriented linguistic processor, which analyzes the texts of NL documents and transforms them into knowledge structures [1,2,9].

A set of meaningful portraits (together with index files) comprise the Knowledge Base (KB) where various types are provided of semantic search and logical analytical functions by comparison and transformation of knowledge structures. We design the technology which provides the processing in the KB distributed within the net of computers.

A meaningful portrait consists of the elementary fragments, arguments of which are words in the normal form (it is necessary for the search and processing). Each elementary fragment has its unique code, which is written in the form of the number with the sign + and is separated by a slash line. For example, in the fragment FIO("ABU - TARIK"," "," "," "/3+) the sign “3+” is its code (but “3-” is the reference to it). Fragments DOC_(22, TERR_DOC.TXT”, “SUMMARY; ” /0+) 0-(ENG) indicate that the meaningful portrait is built on the basis of the English-language text of document with number 66 of the file of TERR_DOC.TXT”, which was processed as the summary of the incidents (linguistic knowledge depends on this). The following fragments present date DATE_(…/1+), criminal group CRIM_GROUP(…/2+), person’s surname (name and patronymic) FIO(… /3+) and so on. The signs “1+”, “1-” and “2+”, “2-” and “3+”, “3-”, … are the codes of the fragments, corresponding to the NE. With the aid of the codes the connections and relations of NE are assigned. Actions are represented in the form of fragments of the type DESTROY(ALIEN,3-/4+) 4-(66,ACT_), where it is represented as “alien person (FIO with code “3+”), are destroyed”. Fragment 4-(66, ACT_) indicates that the first fragment DESTROY(…./4+) presents the action and relates to the document with the number 66. Fragments PLACE_(CHECHEN,REPUBLIC/5+) WHERE(4-,5-) indicate the place of this action (WHERE). Fragments ORG_(…/6+) INFORM(6-/7+) 7-(66,ACT_) represent that “organization … was informed”.

The fragments PREDL_(...), which correspond to the sentences play the special role. They are filled up with the words, which did not enter into the named entities (in this example they are absent), or with the codes of entities themselves. To these fragments the indicators of their position in the text are added. For example, the fragment SENTENSE_(66,12-,16-/20+) 20-(5,254,471) represents the fact that the entities with codes “12-” (corresponding to the action “inform”), “16-” (corresponding the action “destroy”) are located in the sentence, which begins from the 5th line of the text of the document and they occupy the place from the 254-th to the 471-th byte. These means of positioning are necessary for the work of the reverse linguistic processor.

A set of meaningful portraits of documents are organized in the Knowledge Base. Logical reference is provided with the aid of the rules IF… THEN (productions) of the language DECL, which are the basis for decision of logical-analytical tasks.

# 4 Semantic-oriented linguistic processor

Semantics-oriented linguistic processor consist of the following components [9,14].

## 4.1 The component of lexical and morphological analysis (LMA)

It extracts words and sentences from the text, performs lemmatization of words (normal form establishment) and constructs the semantic network presenting the space structure of text (SpST), which reflects the sequence of words, their basic features, beginnings of sentences and the presence of space character lines. The component LMA uses a two-level general ontology and a special collection of subject dictionaries (the dictionary of countries, regions of Russia, names, forms of weapons, and other items specific for the supported domains). The component performs semantic grouping of the words and assigns them additional semantic attributes [10].

## 4.2 The component of syntactic-semantic analysis (SSA)

It converts one semantic network (SN) into another which represents the semantic structure of text (SemST), i.e., the relevant semantic entities and their connections [3,8,9]. The SemST is called the meaningful portrait of document. It comprises knowledge structures of the knowledge base which serves the basis for implementing different forms of semantic search : the search by features and connections, the search for the entities connected at different levels, the search for similar persons and incidents, the search by distinctive characteristics (with the use of ontology).

The component SSA is controlled by the linguistic knowledge (LK), which determines the process of text analysis. LK includes the special contextual rules which ensure the high degree of selectivity with the extraction of entities and connections The functions of this component are the following:

- Extraction of entities from the flow of NL documents: persons, organizations, actions, their place and time, and many other relevant types of entities.
- The establishment of connections between entities. For example, persons are connected with organizations (PLACE_OF_WORK), by addresses (LIVES, REGISTERED). Or figurants of criminal events are connected with such entities as the type of weapon, drugs (TO HAVE).
- The analysis of finite and nonfinite verbal forms with the identification of the participation of entities in the appropriate actions. For example, one figurant gave the drugs to another figurant, and this is the fact linking them.
- The establishment of the connections of actions with the place and time (where and when some action or event occurred).
- The analysis of the reason-consequence and temporary connections between actions and events.

## 4.3 Expert system component (ES)

On the basis of semantic networks the new knowledge pieces are constructed in the form of additional fragments (ESN). For example, the component

ES extracts the field of a person's activity (in accordance with the assigned classifier) from the text of resume for each autobiography. The person's experience in his field is evaluated. The correlation of a criminal incident to the specific type is accomplished with the analysis of the criminal actions of ES: the following facts are revealed - the nature of crime, the method of its accomplishment, the instrument of crime, and so forth (in accordance with the classifiers of the criminal police) [3,12,13].

# 5 The base of linguistic and expert knowledge (KB)

It contains the rules of the text analysis and expert solutions in the internal representation. They determine the work of the linguistic processor. Our logicalanalytical systems have several such bases, which are activated depending on subject areas and user tasks.

## 5.1 Linguistic knowledge

Linguistic knowledge has same structures for various language that give possibilities to tune the processor LP on the text collection in this language, for example, Russian and English. Linguistic knowledge is written in language SSN which has declarative structures. It provides the tuning to new subject field and language for comparatively short time. Procedures part of LP is not changed (excluding blocks of lexical morphological analysis).

## 5.2. Terminological analysis and transformations

Terminological analysis has as a goal - synonymous transformations, the interpretation of abbreviations and the selection of terms. The fragments of the following form are used for this:

TERMIN (<resulting word>,<word1>,<word2>) or
TERMIN (<resulting word>,<word1>,<word2>,<word3>),

where <word1>,… may be normalized word (in canonical form), or sign, or AND-OR graphs. These graphs are represented as fragments STR_OR (...), where facultative words or their signs are on argument places. For example, fragment

TERMIN (UNEMPLOY,NO,WORK)

indicates the conversion: NO WORK - > UNEMPLOY. Another example: fragments

TERMIN (MO,MOSCOW,3+) STR_OR(REGION,REG.,DISTRICT,…/3-)

perform many conversions: MOSCOW REGION - > MO, MOSCOW REG. - > MO, MOSCOW DISTRICT - > MO… For these fragments the permissible context can be assigned (words, which can stand to the left and to the right). The inadmissible context can be also indicated - word or their signs, which there must not be to the left or to the right. As a result it is possible to extract terms and word combinations, whose values depend on context.

Synonyms are presented by fragments:

SYNON (<resulting word>,<word1>,<word2).

For example, SYNON (UKRAINIAN,HOHOL) - word HOHOL (specific name of Ukrainian persons) must be substituted on UKRAINIAN. Many synonyms have conditional nature. The permissible or inadmissible context is indicated for them. For example, in the case given above replacements for the words - surnames, the nicknames, names streets and others are not admitted.

The ontology is presented by fragments of ESN with name SUB (class – subclass), NEAR (nearness of meaning) and OR_OR (separate "or"). For example:

SUB(MAN,TERRORIST)
SUB(TERRORIST,SEPARATIST)
SUB(TERRORIST,REBEL)
SUB(TERRORIST,INSURGENT)
SUB(TERRORIST,MERCENARY) . . .
NEAR(ALCOGOL,DRUNK,TIPSY,VODKA) . . .
OR_OR(MALE,FEMALE,CHILD) . . .

## 5.2. Contextual rules

The block of syntactical-semantic analysis on the basis of context extracts the named entities (NE) and the connected information (for the persons their birthday, sex, address and other) [2,15]. For this contextual rules are used. Syntactical-semantic analysis is necessary for the extraction of addresses, attributes of machines, organizations and other. Usually the entities are the collections of words, which grammatically are not connected. For example, an address can be considered as the collection of letter combinations st. (street), h. (house),.:., words with the capital letter and the numbers. Each such collection can have its boundaries and inadmissible components. For example, in the addresses there can be no verbs and so on. The extraction of such word collections (descriptions of NE) is based on the use of contextual rules of the following form:

CONTEXT (<word1>,<word2>,<word3>,…) - > <resulting fragment>

where <word1>,… - are the normalized word or sign or AND-OR graphs. For every rule a special fragment indicates the position to begin application, and also a permissible or an inadmissible context. This ensures the differentiated application of rules. These rules analyze word group, which describe any entities, and substitute them (in case of application) by one word, with which the resulting fragment is connected.

Contextual rules are applied in the determined sequence. At first they extracted the separated entities, then their properties, word combination, and finally, verbal forms. In process of rules application the meaningful portrait of document will be build.

For example, let us examine rule GG~1:

MUSTBE (GG~1,1) STR_OR(ADJ,PRON/2+)

CONTEXT (2-, NOUN/GG~1) P_P (GG~1,3+) WORD_C (1,2/3-)
3-(2,MORF) NOTBE (GG~1,2,LETT)

This rule provides the conversions:
ADJECTIVE + NOUN → <word combination> and
PRONOUN + NOUN → <word combination>.

The fragment MUSTBE indicates that application of rule GG~1 must begin from the first position, i.e., from search for words with the signs ADJECTIVE (ADJ) and PRONOUN (PRON), since they are fewer than NOUN (NOUN). Fragment P_P separates the left side from the right side (→ ), and WORD_C - indicates that the words on the first and second positions must be united into the combination of words, which subsequently will be considered as one word with the morphological properties of the second word. The fragment NOTBE indicates that in the second position there cannot be the separate letters (sign LETT).

This is an example of a simplest rule. The fragments which indicate the context are added to such rules; these help to restore the full name of an entity, for example, on the basis of pronouns or brief descriptions (by the name the surname is restored, if they were somewhere mentioned together), etc.

Each contextual rule is a semantic network (ESN). All linguistic knowledge is written in the ESN language. The application of rules is provided by the productions of the DEKL language. These productions are organized as programs, which play the role of the empty linguistic shell, which supports the language of the record of linguistic knowledge - ESN. As it is shown by experience, this shell can be tuned to different subject fields and languages. In this way different linguistic processors are designed.

## 5.3. Application of the rules

Application of contextual rules is fulfilled in the strictly defined sequence - each at their level. For example, in system “Criminal” the linguistic processor at first extracts the following named entities - police department, the policemen and others. They can contain surnames, names. It is necessary to facilitate the subsequent analysis. Otherwise the words, which compose these entities, can be captured by other rules and create noise. Further figurants are extracted and so on. The set of rules is introduced for this. Some begin their application from the search of names, surnames (MUSTBE), others - from the search for the birthday, or from the initials. In this way we minimize losses in cases when the block of morphological analysis does not give the necessary data for some words. Then the word combinations are analyzed, and finally, verbal forms. In the process of application of these rules the semantic network (meaningful portrait of document) is built. An example of the levels, which determine the order of the rule application, is given below.

{== levels ==}
LEVEL (LEVEL_E1, LEVEL_E2, LEVEL_E3, LEVEL_E4,…)
LEVEL_E1 (CATALOG) {= extraction of word combinations from the catalogs =}
LEVEL_E2 (MIL~~1, ST~~1) {= extraction of police departments,… =}
LEVEL_E3 (FF~~1, FF~~2) {= extraction of figurants =}
{== grammatical analysis, the extraction of word combinations =}
{== AA~~… - uniform terms, GG~~… - words combination ==}
LEVEL_E4 (AA~~1, AA~~3, AA~~4, GG~~1, GG~~2,…)

{= GG~~1: word combination ADJECTIVE – ADJ or PRONOUNCE - NOUN=}
MUSTBE (GG~~1,1) STR_OR (ADJ, PRON/2+) CONTEXT (2-, OBJ/GG~~1)
P_P (GG~~1,3-) WORD_C (1,2/3+) 3- (2, MORF) NOTBE (GG~~1,2, LETT)
. . .

In the curly braces the commentaries are given. It is an example of the rule GG~1, which reveals word combinations with signs ADJ or PRON and OBJ (i.e. NOUN etc.). The system has a full set of contextual rules, which provide the complete analysis of sentences and construction of meaningful portraits of documents. But in contrast to the standard grammars our LP provides the extraction of all significant (information) entities, including those in which the words are not connected grammatically, for example, addresses, machines with the indication of their numbers and so on. The described processor (LP) is semantics-oriented, because it provides the extraction of entities and supports various kinds of connections between them. These are semantic components. Such LP found their use in the systems of a new class – “Analytic”, “Criminal”, “AntiTerror” and other.

# 6 Conclusion

The described semantic-oriented linguistic processors have been used for construction of intellectual analytical systems: “Criminal”, “Analytic”, “AntiTerror”, “Monuments” and others. The distinctive features of these systems are characterized by universality of the processor. It provides automatic extraction of knowledge structures from texts in various language. Now they are Russian and English. The processor can be quickly tuned to other European languages. The result of the processor operation is the Knowledge Base which has common structure for all language and which is used for realization of logical-analytical functions. The ESN apparatus provides powerful representational possibilities for describing all levels of natural language, including the level of deep semantic structures, and cross-lingual correspondences.

The implemented linguistic processors were created on the basis of this approach which made it possible to manufacture design solutions for the basic problems of extracting meaningful knowledge from the texts in natural languages.

References

[1] Kuznetsov, I.P. Elena B. Kozerenko, Mikhail M. Charnine. Technological peculiarity of knowledge extraction for logical-analytical systems // Proceedings of ICAI'12, WORLDCOMP'12, July 18-21, 2012, Las Vegas, Nevada, USA. - CRSEA Press, USA, 2012.

[2] Kuznetsov I.P. Matskrvich A.G. Semantic oriented systems controlled by knowledge base // University of communications and informatics, Moscow, 2007,173 p.

[3] Kuznetsov I.P. Methods of Processing Reports with the Extraction of Figurants and Events Features // In "Computational Linguistics and its Applications", Vol.2, Tarusa, 1999.

[4] Kuznetsov I.P. Semantic Representations // Moscow: "Nauka", 1986. 290p.

[5] Kozerenko, E.B. Multilingual Processors: a Unified Approach to Semantic and Syntactic Knowledge Presentation. In Proceedings of the International Conference on Artificial Intelligence IC-AI'2001 25-28, 2001. CSREA Press, 2001, pp.1277-1282.

[6] Byrd, R. and Ravin, Y. Identifying and Extracting Relations in Text // 4th International Conference on Applications of Natural Language to Information Systems (NLDB). Klagenfurt, Austria, 1999.

[7] Kuznetsov I.P. Natural Language Texts Processing Employing the Knowledge Base Technology // Sistemy i Sredstva Informatiki, Vol.13, Moscow: Nauka, 2003, pp. 241-250.

[8] Kuznetsov I.P., Matskevich A.G. The English Language Version of Automatic Extraction of Meaningful Information from Natural Language Texts // Proceedings of the Dialog-2005 International Conference "Computational Linguistics and Intelligent Technologies", Zvenigorod, 2005pp. 303-311.

[9] Kuznetsov, I., Kozerenko, E. The system for extracting semantic information from natural language texts // Proceeding of International Conference on Machine Learning. MLMTA-03, Las Vegas US, 23-26 June 2003, p. 75-80.

.

[10] Somin N.V., Solovyova N.S., Charnine M.M The System for Morphological Analysis: the Experience of Employment and Modification // Sistemy i Sredstva Informatiki, Vol. 15 Moscow: Nauka, 2005, pp. 20-30.

[11] Cunningham, H. Automatic Information Extraction // Encyclopedia of Language and Linguistics, 2cnd ed. Elsevier, 2005.

[12] Kuznetsov I.P., Matskevich A.G. Semantics Oriented Linguistic Processor for Automatic Formalization of Autobiographical Data // Proceedings of the Dialog-2006 International Conference "Computational Linguistics and Intelligent Technologies", Bekasovo, 2006, pp. 317-322.

[13] Web site "Knowledge extraction for Analytical Systems": http://Ipiranlogos.com/english/

[14] Kuznetsov I.P., Kozerenko E.B., Matskevich A.G. Deep and Shallow Semantic presentations in Intelligent Fact Extractors // Proceedings of ICAI'2010 Las Vegas, USA, June 14-17, 2010, CRSEA Press, 2010.

[15] Kuznetsov, I.P., Kozerenko E.B. Semantic Approach to Explicit and Implicit Knowledge Extraction // Proceedings of ICAI'11, WORLDCOMP'11, July 18-21, 2011, Las Vegas, Nevada, USA. - CRSEA Press, USA, 2011.

[16] Clark P., P. Harrison, and J. Thompson. A Knowledge-Driven Approach to Text Meaning Processing // Proceedings of the HLT-NAACL 2003 Workshop on Text Meaning, 2007. P. 1–6.

[17] Gildea D. and M. Palmer. The necessity of syntactic parsing for predicate argument recognition. In Proceedings of the 40th Annual Conference of the Association for Computational Linguistics (ACL-02), Philadelphia, PA, 2002. P. 239–246.

[18] Pasca M. and B. Van Durme. What You Seek is What You Get: Extraction of Class Attributes from Query Logs // Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI-07), 2007. P. 2832–2837.

[19] Punyakanok V., D. Roth, and W. tau Yih. The Importance of Syntactic Parsing and Inference in Semantic Role Labeling // Computational Linguistics 34(2), 2008. P. 257–287.