

# ВЫЯВЛЕНИЕ РОЛЕВЫХ ФУНКЦИЙ ЛИЦ НА ОСНОВЕ СТРУКТУР ЗНАНИЙ

**И. П. Кузнецов** (igor-kuz@mtu-net.ru)

ИПИ РАН Вавилова, Москва, Россия

Рассматривается семантико-ориентированный лингвистический процессор, извлекающий из текстов естественного языка информационные объекты, их свойства и связи и формирующий на этой основе структуры знаний. Одно из направлений развития таких процессоров связано с выявлением имплицитной информации, которая рассматривается в узком плане — как выявление новых свойств объектов, заданных в неявном виде. Предлагается методика такого выявления, основанная на анализе структур знаний. В качестве примера рассматривается выявление ролевых функций фигурантов на базе их описаний в сводках происшествий.

**Ключевые слова:** ролевые функции, структуры знаний, лингвистический процессор, фигуранты, происшествия, сводки происшествий.

## IDENTIFYING ROLE FUNCTIONS OF PEOPLE ON THE BASIS OF KNOWLEDGE STRUCTURES

**I. P. Kuznetsov** (igor-kuz@mtu-net.ru)

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

The linguistic processor which extracts knowledge structures (information objects and their links) from natural language texts is considered. The development of the processor is connected with extracting implicit information, e. g. role functions of people. The proposed extraction methods are based on the analysis of knowledge structures. The methods are used for identification of role functions of people involved in criminal cases reported in law texts.

**Key words:** linguistic processor, role functions, knowledge structures, figurants, criminal cases, reports.

## Introduction

One of the primary tasks in the area of cognitive technologies is the automatic extraction of knowledge from natural language (NL) texts. It is a complex problem connected with developing linguistic processors which perform automatic formalization of texts, i. e. mapping the texts into formal models or knowledge structures. It should be noted that a lot of relevant information in NL texts is presented in a concealed form. This information is called implicit. An example is the task of assigning certain features to persons on the basis of actions performed by them. In the subject area of "Criminology" it is assigning such features as "victim", "suspect", etc. to persons. The methods of implicit information extraction should be considered in the context of the knowledge extraction task, and it is conditioned by the specific features of the linguistic processor. The given paper describes these methods within the framework of the object-oriented linguistic processor developed at the Institute for Informatics Problems of the Russian Academy of Sciences (IIPRAS).

### 1. The object-oriented linguistic processor

The research direction connected with unstructured NL texts processing has been developing for 20 years at the IIPRAS for particular application areas and specific user tasks [1,2]. One should consider that the large category of users have the specific official responsibilities, and respectively, constant interests. Completely concrete information is necessary for them. For example, a criminal inspector seeks to extract information on important figurants, their places of residence, telephones, criminal events, dates and other such facts [3]; a personnel manager is interested in the organizations, when and where a person worked and in what position [4]. Other people try to fish out from the media the information about the countries, important persons, catastrophes, places of interest and historical monuments [5]. We call this concrete information interesting for a user *information objects*. Objects are distinguished by their types.

Let us note that the connections between the objects, which interest users, can have the high degree of variety. For example, not only the connection of the persons with their information from their questionnaires or the objects can present interest, but also the actions or the events, in which these persons participate. Such events are attached to the time and the place. Moreover, some events can be a component part of others. They can be connected with cause-effect and temporary relations. For the number of problems similar connections play an important role. They also must be revealed and processed. Therefore one should consider that events are also information objects, interconnected and connected with other information objects. Complex structures appear. For their representation within the framework of the projects of IIPRAS (Russian Academy of Science) the language of the extended semantic networks (ESN) has been developed, while for the processing the production rules the language DEKL [6] has been implemented.

The ESN networks are represented in the form of special graphs [6]. In the formal record they are the extension of the predicate logic language. ESN consist

of elementary fragments, each of which has its unique code (see Section 3), which can stand at the argument places of other fragments, and provide great possibilities for the representation of knowledge structures. The language DEKL is designed for the transformation of such structures. The problem of extracting knowledge from natural language texts is considered from the point of view of developing information objects and connections with the construction of the knowledge structures on the basis of which the solution of user problems is achieved. For this within the framework of the IIPRAS projects the *object-oriented* linguistic processor (LP) converting the natural language (NL) texts to the knowledge structures is developed and constantly updated. The processor LP achieves a deep NL text analysis with bringing of synonymous groups to one form, development of objects and their properties, identification of objects, elimination of ambiguities, development and unification of various forms, which present events or actions (including forms with the verbal noun, participial and verbal-adverbial constructions), which are connected with the time and the place [2–6]. As a result the structures of knowledge in the ESN formalism are created.

The linguistic processor (LP) is realized by means of the language DEKL and is controlled by the linguistic knowledge (LK) in the form of object dictionaries, means of parametric tuning, and also the rules of extracting objects and connections [2,4,5,6]. With the aid of LK the tuning of LP to the appropriate categories of users and text corpora is accomplished. Concrete realization appears as a result. Thus, the paper deals with the means of constructing a class of processors with powerful mechanisms for their tuning and updating. Further development of such processors (LP) is connected with the development of implicit information [7], which we will consider in a narrow plan, i. e. as the addition of the structures of knowledge by the new information, which is absent or assigned implicitly. In this article the procedure of this development is proposed, which consists in the use of LP for mapping NL texts onto the structures of knowledge (ESN) and the use of the means of logical-analytical processing (productions of the language DEKL) for the creation of new information.

Advantages and deficiencies of the proposed procedure will be examined on specific objectives from the area of “criminology”, that is the role functions establishment for the persons (participants) on the basis of the acts performed by them or due to the participation in some specific events. We consider the problem of assignment of properties to the persons (basing on their participation in the acts of different kinds) — “the suffered”, “the suspect” and others, if an explicit description of such properties is absent from the text. For example, if it is said in the text “suffered Ivanov I. I.”, then another task appears, i. e. extraction of some property in the process of linguistic analysis and forming of the corresponding fragments in the knowledge structure. In this article the discussion will deal with LP, customized for the Russian language texts (NL), although the possibilities of LP are wider. There is a sufficient test of tuning LP to the English language texts [9,10].

## 2. The choice of method

The task of the role functions establishment for the information objects is a special case of the more general task, connected with the estimation of objects according

to their descriptions in the NL texts, for example, with the estimation of the stability of enterprise (according to the information from the Internet), by featuring political figures (positive or negative depending on the statements in the press), by the estimation of the role functions quality of product (basing on the statements of users) and so forth. Quite frequently, it is not said directly whether something is bad, or good. As a rule, in NL texts the events are described, the situations, in which one or other information object participated. On the basis of them the estimation is done, which is often represented in the form of a new (generated) property of object.

For the solution of this problem different methods are used [11–15]. The most common one is the method of the new properties of objects development by using the syntactical-semantic forms. For example:

*<what-medicine> caused allergy in <who-human organism>...,  
< what-medicine > has side effects ...  
<who-person> made scandal... and so forth.*

The application of such forms to the NL texts consists in the search for “estimating” or “characterizing” words (of type “scandal”) or for word combinations of the type “caused allergy” (“it can cause allergy”), it “has side effects” (“side-line actions”), “to make scandal” (“to brawl”)... And then the environment is analyzed, i. e., the words, which stand to the left and to the right, their semantic classes (objects are recognized by them) and case forms. Estimations of information objects as a result are given. By the first two forms the “quality of medicines” is estimated, while by the latter it is recognized that a man performed “hooligan actions” or that he is “suspected”. It is known that in NL many versions are possible for expressing the same idea — with the aid of different syntactic constructions, verbal groups, forms and so forth. Therefore the number of estimating word combinations will be sufficiently large. Moreover, the application of such forms requires different forms of analysis — morphological (in order to reduce different word forms to one form), syntactic (the trees are built of the selection of sentences in order to isolate the connected components and to find place for the estimated words) and semantic (in order to extract the objects, which are evaluated). The use of syntactical-semantic forms is connected with certain difficulties caused by special NL features: by the presence in texts of participial, verbal-adverbial constructions, different explanations, facultative components (time, place, purpose), anaphoric references and other language structures. As a result, information objects are frequently disconnected from the estimated words. Hence — the significant losses, which influence the quality of estimation.

**Example 1** (*the text is taken from the summaries of incidents of the City Office of Home Affairs, Moscow*):

*... Gorelov Peter Sergeevich, 01.03.76 yr/bir, liv: c. Moscow, st. Young Leninists, h.71-6-12, does not work, 01.02.1998 yr. at 4.30 in his house out of hooligan motives in the state of alcoholic intoxication made scandal and broke window glass in the apartment of Litvinova Galina Ivanovna, 20.07.1961 yr/bir,...*

**Пример 1** (текст взят из сводок происшествий ГУВД г. Москвы):

*... Горелов Петр Сергеевич, 01.03.76 г/р, прож: г.Москва, ул.Юных Ленинцев, д.71-6-12, не работает, 01.02.1998 г. в 14.30 у своего дома из хулиганских побуждений в состоянии алкогольного опьянения учинил скандал и разбил оконное стекло в квартире Литвиновой Галины Ивановны, 20.07.1961 г/р, ...*

In this example the estimating (characteristic) words are “made scandal” and “broke the window glass”, they are located at a significant distance from the estimated person — “Gorelov Peter Sergeyevich”. This limits the possibilities of applying the forms. It is required that the initial extraction of components, which must not be considered in the forms: the years of birth, addresses, specific properties (“he does not work”, “in the state of alcoholic intoxication”), time, place and others, which requires sufficiently deep text analysis with the extraction of objects, their properties and attributes. In connection with the aforesaid, another more promising method is represented — when evaluation is accomplished at the level of knowledge structures. For their construction the objective-oriented LP is used producing the structures of knowledge in which the objects are directly connected with the events and the actions and excluding the above mentioned losses. For the development of implicit information (role functions of objects) the rules of the DEKL language are used which analyze the structures of knowledge (ESN) and form new properties of objects. In this case the structure of knowledge does not change, but it is only supplemented by new (useful) fragments.

### 3. Meaningful portraits of documents

Within the framework of the proposed procedure the development of the role functions of objects (implicit information) is achieved at the level of the structures of knowledge, called the meaningful portraits of documents (SS-documents). Let us examine how such structures appear in the ESN formalism [2,3,6].

**Example 2 (translation of the Russian text given below).** *Text N22 is taken from the summaries of incidents of the City Office of Home Affairs, Moscow: 01.02.98 yr. 16–30 to the Home Office applied citizen Mitrofanov Victor Mikhailovich, 1955 yr. bir, liv.: Bohr Highway 38–211, n/w. he stated that 01.02.98 yr. at 10-00 in house 3 at St. Fedosino the unknowns being found in the drunk state made scandal, they expressed themselves by unquotable swearing, they set dog. As a result of what Mitrofanov applied to trauma care center, where the diagnosis was set: the bite of foot.*

**Пример 2.** Текст документа (с номером 22) взят из сводок ГУВД:

*01.02.98 г. в 16-30 в ОВД обратился гр-н Митрофанов Виктор Михайлович, 1955 г.р., прож.: Боровское шоссе 38-211, н/р. Он заявил, что 01.02.98 г. в 10-00 у д.3 по ул. Федосьино неизвестные находясь в пьяном виде учинили скандал, выразились нецензурной бранью, натравили собаку. В результате*

*чего Митрофанов обратился в травмпункт, где был поставлен диагноз: укус ноги.*

The objective-oriented LP performs the deep analysis of the text and automatically builds its meaningful portrait (SS- document, transliterated):

DOC\_(22, "1-02-98", "SUMMARY; " /0+) 0 (RUS)  
OVD\_(OVD/1+)  
FIO(MITROFANOV], VICTOR, MIKHAYLOVICH, 1955/2+) UNEMPLOYED (2-/3+) 3-  
(22, PROP)  
ADR\_(Borovskiy, Sh., 38,211/4+)  
PROZH. (it is 2nd, 4)  
ADR\_(UL, FEDOSINO, HOUSE, 3/5+)  
FIO (" ", " ", " ", " ", NESKOLKO/6+)  
UNKNOWN (6)  
DRUNK (6-/7+) 7 (2, PROP\_)  
SCANDAL (6, PYANYY/8+)  
IS EIGHTH (22, ACT\_)  
TO REPORT (IT IS 2ND, 8-/9+) 9 (22, ACT\_)  
DATA\_(1998,02, ~01, " 10-00" /10+)  
When (9, 10)  
TO TURN (1, GR- N, 2-/11+) 11- (22, ACT\_)  
DATA\_(1998,02, ~01, " 16-30" /12+)  
When (11-, 12-)  
EXPRESS (6, UNQUOTABLE, [BRAN]/13+) 13- (22, ACT\_)  
TO SET (6, [SOBAKA]/14+) 14 (0, ACT\_)  
TO TURN (IT IS 2ND, IN, [TRAVMPUNKT]/14+) 14 (0, ACT\_)  
TO PLACE (DIAGNOSIS, BITE, [NOGA]/16+) 16 (0, ACT\_)  
PREDL\_(22,11-, 4, 3-, 9, 13-, 14-/17+) 17- (2,15,341)  
PREDL\_(22,15-, 16-/18+) 18- (6,342,448)

For the original Russian text the automatically generated SS- document looks as follows:

ДОК\_(22, "1-02-98", "СВОДКА;" /0+) 0-(RUS)  
ОВД\_(ОВД/1+)  
FIO(МИТРОФАНОВ, ВИКТОР, МИХАЙЛОВИЧ, 1955/2+)  
БЕЗРАБОТНЫЙ(2-/3+) 3-(22, PROP\_)  
АДР\_(БОРОВСКИЙ, Ш., 38, 211/4+)  
ПРОЖ. (2-, 4-)  
АДР\_(УЛ., ФЕДОСЬИНО, ДОМ, 3/5+)  
FIO(" ", " ", " ", " ", НЕСКОЛЬКО/6+)  
НЕИЗВЕСТНЫЙ(6-)  
Пьяный(6-/7+) 7-(2, PROP\_)  
СКАНДАЛ(6-, Пьяный/8+) 8-(22, ACT\_)  
СООБЩИТЬ(2-, 8-/9+) 9-(22, ACT\_)  
ДАТА\_(1998, 02, ~01, "10-00"/10+)  
Когда(9-, 10-)  
ОБРАТИТЬСЯ(1-, ГР-Н, 2-/11+) 11-(22, ACT\_)  
ДАТА\_(1998, 02, ~01, "16-30"/12+)  
Когда(11-, 12-)

ВЫРАЖАТЬСЯ(6-, НЕЦЕНЗУРНЫЙ, БРАНЬ/13+) 13-(22, ACT\_)  
 НАТРАВИТЬ(6-, СОБАКА/14+) 14-(0, ACT\_)  
 ОБРАТИТЬСЯ(2-, В, ТРАВМПУНКТ/14+) 14-(0, ACT\_) ПОСТАВИТЬ(ДИАГНОЗ, УКУС, НО  
 ГА/16+) 16-(0, ACT\_) ПРЕДЛ\_(22, 11-, 4-, 3-, 9-, 13-, 14-/17+) 17-(2, 15, 341)  
 ПРЕДЛ\_(22, 15-, 16-/18+) 18-(6, 342, 448)

A meaningful portrait consists of the elementary fragments, arguments of which are words in the normal form (necessarily 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 OVD\_(OVD/1+) the sign 1+ is its code (but 1 is the reference to it). Fragments DOK\_(22, "1-02-98.TXT", "SUMMARY;" /0+) 0 (RUS) indicate that the meaningful portrait is built on the basis of the Russian-language text of document with number 22 of the file of 1-02-98.TXT", which was processed as the summary of the incidents (linguistic knowledge depend on this). The following fragments present police department OVD\_(... /1+), person's surname, name and patronymic FIO (... /2+), person's specific property UNEMPLOYED (2-/3+), address ADR\_ (... /4+) and so forth; the signs 2+, it is 2nd, 3+, 3-,... are the codes of the fragments, with the aid of which their connections and relations are assigned. For example, the fragment PROZH (live) (it is 2nd, 4) represents the relation that the person (represented as FIO with code 2+) lives at the address (fragment [ADR]\_ with code 4+). Actions are represented in the form of fragments of the type SCANDAL (6, PYANYY/8+) it is 8 (22, ACT\_), where it is represented that "person (FIO with code 6+), being drunk, made scandal". With the aid of it is the fragment 8\_(22, ACT\_) indicates that the first fragment is SCANDAL (.../8+) presents the action and relates to the document with the number 22. A similar role is played by the fragments of the type 3-(22, PROP\_), by which the properties are noted. The codes of fragments also serve for the idea of time, scene of action and cases, when one action is included in the composition of another. For example, the fragment TO REPORT (it is 2nd, 8-/9+) represents that the person (code 2+) "reported" (code 9+) about the action (code 8+), i. e., about "made scandal". The following fragments DATA\_(... /10+) when (9, 10) represent the time (DATA\_), which relates (when) to the action "to report". Special role is played by the fragments PREDL\_(...), which correspond to the sentences. They are filled up with the words, which did not enter the information objects (in this example they are absent), or with the codes of objects themselves. To these fragments the indicators of their position in the text are added. For example, the fragment PREDL\_(22,11-, 3-, 9, 13-, 14-/17+) 17- (2,15,341) represents the fact that the objects with codes 11- (corresponding to the action "to turn"), 3- (corresponding to the property "unemployed") and others are located in the sentence, which begins from the 2nd line of the text of the document and they occupy the place from the 15-th to the 310-th byte. These means of positioning are necessary for the work of the reverse linguistic processor (LP).

Analyzing this example, it is possible to make the following conclusions: 1) In SS-document the estimating (characterizing) words occur either in one fragment with the object — SCANDAL (...), or the next one, i. e., the codes of the actions, in which the object participates, are nearby in PREDL\_(... 9, 13-, 14...). In this case the possibility of composite actions is considered. 2) On the actions, represented as SCANDAL (...), it is possible to draw the conclusion that the discussion deals with "that suspected", and

TO REPORT (,) — that the person is “suffered” or “the applicant”. Such conclusions are easily arrived at with the aid of the rules IF... THEN (productions) of the language DEKL, which are the basis for the extraction of role functions. 3) The particular difficulties of dividing the text into the sentences occur (in the old version). The reduction “of n/r” (with the point at the end) was not understood as the end of a sentence. 4) The linguistic processor (LP) correctly identified the pronoun “he”, and also it knew how to reveal the participation of the subject (“*unknowns*”) by the actions “*to be evinced by unquotable swearing*” and “*to set dog*”, which also characterize subject. At the same time the LP could not connect the action “*diagnosis was set*” with the person — “*Mitrofanov...*” (the code is 2-nd). In this case an example proved to be successful. Also the processor LP (with its linguistic knowledge — LK) was developed for the tasks of the criminal police, connected with different forms of the objective searches: the search for similar participants (addresses, and so forth), search according to the connections, precise search for objects, for the search by signs and other identifiers. In this case the analysis of some complex NL forms was not required, i. e. the cases of the enumeration of the objects participating in the uniform actions (they are described by one verb), the enumeration of the actions of one object and others in contrast to the aforesaid, with the extraction of role functions for each object the indication of its participation in each action is required. Hence it follows that with the use of the proposed procedure the more qualitative extraction of role functions is directly connected with the works on improvement of LP in the aspect of the development of objects and their actions. In many instances the numerous errors caused the inaccuracies in SS-document, e. g.: the absence of punctuation marks or their presence (where it was not required), the inappropriate reductions, gaps in the words and many others. The fact is that the documents, entering the summaries of incidents, are composed on the spot by people (militiamen) of different degree of literacy. Hence — the additional noise and loss. Thus, meaningful portraits are the collections of fragments of ESN which represent the sufficiently high level of formalization of NL texts and are convenient for the working — with the aid of the instrument means — DEKL [3]. Besides LP which analyzes texts and builds SS-documents, there is a reverse linguistic processor (LP) which on the basis of the fragments of the SS- document generates the NL texts presented to the user [6].

#### 4. The means of the development of the role functions

As it has already been said, within the framework of the proposed procedure (instead of the application of syntactical-semantic forms to the documents) the rules are used for logical conclusion and transformation of the knowledge structures — the SS- documents, in which there are no morphological features (of type who, whom,...), and the subjects and the objects are distinguished by their arrangement in the fragments of ESN, which present actions. The names of fragments present the nature of actions. Syntactical-semantic forms are transformed into the fragments of ESN which determine conversions and logical conclusion achieved by productions of language DEKL. Such fragments play the role of the logical-semantic shell, which determines conversions and logical conclusion on the basis of SS-documents. After filling of the



shell by ontological-fragmental knowledge (OFK) which consist of the mentioned fragments (ESN), the program is formed, which accomplishes the development of role functions and completion of the SS-document by the appropriate fragments. With this approach it is possible to avoid many difficulties, connected with the design features of NL and the specific character of the use of syntactical-semantic forms. There are many versions of construction of the shells and representation of the corresponding knowledge which are distinguished according to the degree of their generality. Let us examine the version which is at present realized and verified.

**Case 1.** The role functions are determined by the names of actions. In this case for the extraction of objects (participants) which should be assigned properties (role functions), the fragments of the following form are used :

```
INTERPRET (MAN_2, FIO, "suffered") FORMA_CC (MAN_2, CLASS_D4, " ") CLASS_
D4 (TO TURN, TO STATE, TO REPORT, TO PASS AWAY,...)
```

The first fragment INTERPRET (...) means that from the SS- document it is necessary to extract the fragments of the form FIO (...), that correspond to participants, and to analyze the possibility of assigning them the property "suffered". Such participants are conditionally designated as MAN\_2. The second fragment FORMA\_CC (...) specifies the conditions for assigning this property to MAN\_2, determined by the constant CLASS\_D4. In the third fragment CLASS\_D3 (...) the words are given which present actions. It is represented that the words belong to the class CLASS\_D3. If the participant occurs in one of the enumerated actions, then to this participant the property "suffered" is assigned. This participation is revealed via the analysis of the SS-document. If there is a fragment TO TURN (... , it is n-th,...) in it, the argument of which is the code FIO (... /N+), then the fragment N-("suffered") is added that represents the role function of the corresponding participant. Conformably for the SS-document represented in example 2 the analysis will occur as follows. Consecutively the extraction of fragments FIO (...) corresponding to the participants is performed. First FIO (MITROFANOV,... /2+) will be extracted . Its code is 2- is the argument of the fragment TO TURN (1, GR- N, 2-/11+), that presents the action. In connection with this to SS- document the fragment 11- ("suffered") will be added, which via the reverse LP will be transformed into the statement that "Mitrofanov Victor Mikhaylovich is a suffered person". These actions are realized within the framework of the logical-linguistic shell.

**Case 2.** Role functions are determined by the actions and elucidating words. For this the same fragments are used, as in the first case, but during the enumeration of the names of actions the additional fragments which present actions with the possible elucidating words, are introduced:

```
INTERPRET (MAN_1, FIO, "suspect")
FORMA_CC (MAN_1, CLASS_D3, " ")
FRAUD (USER, POKUPATEL/15+)
TO SET (SOBAKA/16+)
TO BE EXPRESSED (UNQUOTABLE, SWEARING, MATERNYY,... /17+)
CLASS_D3 (IS DELAYED, TO BE SOUGHT, ..., 15, 16-, 17-)
```

The given fragments determine actions of the extraction of persons (MAN\_1), by which the property of “suspect” is assigned. For this at the level of the knowledge structures their participation is analyzed in the actions “*is delayed*”, “*to be sought*”, and also in the composite actions: “*to set dog*”, “*to be expressed unquotable...*”, “*to be expressed by swearing...*” and others. In example 2 the code of fragment FIO (" " , " " , " " , " " , NESKOLKO]6+), that represents the unknown persons is the argument of the fragment TO SET (6, SOBAKA/14+), representing action “*to set*” with the elucidating word “*dog*” — “*sobaka*”. Therefore the fragment 6 is added (“*suspect*”), that represents that “*the unknown persons are suspected*”, and through the reverse LP the explanation to this conclusion is offered, see below. A similar conclusion will be made on the basis of the fragment TO BE EXPRESSED (6, UNQUOTABLE, BRAN/13+), but with other explanations.

**Case 3.** The actions determine the role functions of several persons. For this (additionally to the fragments INTERPRET) the fragments are added: CLASS\_D1 (TO STRIKE, TO BEAT UP,...) FORMA\_CC (MAN\_1, CLASS\_D1, MAN\_2), where FORMA\_CC (...) indicates the need of the search of two persons — “*suspect*” and “*suffered*” (MAN\_1 and MAN\_2), that participate in one action, which are mentioned in the fragment CLASS\_D1 (...). For example, “*certain person struck another...*”. In the appropriate fragment TO STRIKE (...) the code FIO (...) that corresponds to the first person will stand in front of the second. The given fragments of ESN compose the knowledge OFK which are constantly supplemented — due to the filling of classes by the new words-actions and with the elucidating words. The process of filling is sufficiently simple. If role function is not revealed, then it is necessary to look in the SS-document in which the action of one or another participant (by the text its role is easily determined) occurs. Further, the corresponding constants are located, by which are supplemented the classes of knowledge OFK. Subsequently it is intended to automate the process of completing the knowledge OFK as follows. In the text the words, which determine role functions, are noted. Further, in the formed SS-document the corresponding constants which supplement knowledge OFK are located.

## 5. Explanation of the results

The explanation of results is accomplished through the reverse LP which on the basis of SS- document and additional fragments builds texts in natural language which are displayed to the user. The reverse LP through the codes of the fragments which correspond to object and actions, finds the sentence (PREDL) and its location in the text of a document. Through the arguments of these fragments (the words in the normal form) the processor finds the components of the sentences in which the mentioned object and actions are described. These components are converted into the form suitable for the delivery to the user. The fact is that many components are transformed depending on the context. For example, “... *he threatened Petrov I. I....*” — “... *ugrozhal Petrovu I. I....*”, where during the delivery FIO “*Petrovu*” should be transformed into “*Petrov I. I.*” Further the description of the object is delivered, its generated property and actions which explain this property — role function.

**Example 3.** with the use of the above given knowledge OFK of the SS — document of example 1 the role functions will be generated which with the aid of the reverse LP will be given out to user in the following form:

unknowns — suspected,  
since — unknowns, being in the drunk state made scandal,  
they expressing themselves by unquotable swearing, they set dog  
Mitrofanov Eugene Mikhaylovich, 1953 r. — suffered,  
since — applied to OVD citizen Mitrofanov Eugene Mikhaylovich, 1953 yr. birth,  
since Mitrofanov applied to trauma care centre.

It should be noted that all the actions considered in the paper connected with various cases of role functions establishment and explanations of the results are implemented within the framework of the logical semantic environment written in the logical programming language DEKL. Since the DEKL language is oriented at the processing of knowledge structures (represented in the form of the extended semantic networks — ESN) and since it features the generalized production rules [6], the program code in the DEKL language is very simple and concise: it comprises 16 productions and about 4 Kbyte of text.

## Conclusions

The proposed procedure of the role functions extraction centered at the analysis of knowledge structures is sufficiently promising from the point of view of the knowledge bases technology development. The current task is to improve its performance for the documents comprising enumeration of the type: 1. *Ivanov I. I.*... 2. *Petrov A. A.*... 3.... and further follows the continuation, which describes their acts, for example, “*were subjected to detention by...*” or “*who performed...*”. The recognition of such cases requires further upgrade of the linguistic processor (LP) software. The quality of analysis is lowered by the breaks in significant words of the type “*Iva nov*” or “*Iva-nov*”, which are typical for the summaries of incidents. The methods were tested on the basis of the summaries of incidents which contain about three thousand documents (each document consists of 10–80 lines). In the case of the summaries processing the documents with the mentioned enumerations (there were about 10% of them) were withdrawn and in the remained texts the gaps in the words were removed. At the current moment the program which realizes the proposed procedure gave about 80% of correct recognition of role functions, and about 65% of complete explanations with the indication of all acts. But these numbers rapidly change for the better due to the means (the LK and OFK knowledge) of tuning the LP to special features of the subject area texts. For this not much time is required. Let us note that tuning itself to the extraction of the role functions of persons from the mentioned summaries (with reaching the indicated percentages), required about two weeks of the work of one person. The development and fixing of the shell itself took about four days. The subsequent development is connected with the improvement and the tuning of LP to the work with complex NL forms. At present the extraction of actions is interfered with causal word combinations of the type “*out*

of the hooligan motives”, “owing to the hostile relations” and so forth, which at present are introduced into the system. Difficulties appear with the transfer of the subject of action to other actions to which the subject is not assigned explicitly, but its presence is implied.

The second direction of research and development is connected with the extension of the shell features to the solution of other problems connected with the estimation of objects depending on the nature of statements about them in the texts of description. Within the framework of the studies conducted it is also intended to tune the shell to the work with the English language texts. Since the meaningful portraits of the English language and Russian language texts have the identical structure (SS-documents), this tuning cannot be labor-consuming.

## References

1. Asher N., Lascarides A. 2003. Logics of Conversation.
2. Banko M., Cafarella M., Soderland S., Broadhead M., Etzioni O. 2007. Open Information Extraction from the Web. Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI-07) : 2670–2676.
3. Clark P., Harrison P., Thompson J. 2007. A Knowledge-Driven Approach to Text Meaning Processing. Proceedings of the HLT-NAACL 2003 Workshop on Text Meaning : 1–6.
4. Gildea D., M. Palmer. 2002. The Necessity of Syntactic Parsing for Predicate Argument Recognition. Proceedings of the 40th Annual Conference of the Association for Computational Linguistics (ACL-02) : 239–246.
5. Kuznetsov I. P. 1986. Semantic Ideas [Semanticheskie Predstavleniia].
6. Kuznetsov I., Kozerenko E. 2003. The System for Extracting Semantic Information from Natural Language Texts. Proceeding of International Conference on Machine Learning. MLMTA-03 : 75–80.
7. Kuznetsov I. P. 1999. Methods of Reports Editing with Figurant and Incidents Characteristics Selection [Metody Obrabotki Svodok s Vydeleniem Osobennostei Figurantov i Prosshestvii]. Komp’iuternaia Lingvistika i Intellektual’nye Tekhnologii: Trudy Mezhdunarodnoi Konferentsii “Dialog 1999” (Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference “Dialog 1999”).
8. Kuznetsov I. P., Matskevich A. G. 2006. Semantically Oriented Linguistic Pre-processor for Automatic Formalization of Autobiographical Data [Semantiko-Orientirovannyi Lingvisticheskii Protsessor dla Avtomaticheskoi Formalizatsii Avtobiograficheskikh Dannyykh]. Komp’iuternaia Lingvistika i Intellektual’nye Tekhnologii: Trudy Mezhdunarodnoi Konferentsii “Dialog 2006” (Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference “Dialog 2006”) : 317–322.
9. Kuznetsov I. P., Efimov D. A. 2008. The Peculiarities of Knowledge Extraction by Semantically Oriented Linguistic Processor Semantix [Osobennosti Izvlecheniia Znaniy Semantiko-Orientirovannym Lingvisticheskim Protsessorom Semantix]. Komp’iuternaia Lingvistika i Intellektual’nye Tekhnologii: Trudy

- Mezhdunarodnoi Konferentsii "Dialog 2008" (Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference "Dialog 2008"), 7 (14) : 281–291.
10. *Kuznetsov I. P., Matskevich A. G.* 2007. Semantically Oriented Systems with Knowledge Bases [Semantiko-Orientirovannye Sistemy na Osnovanii Baz Znani].
  11. *Kuznetsov I. P., Somin N. V.* 2008. The Facilitates of Semantically Oriented Linguistic Processor Adjustment for Search and Finding Out the Objects [Sredstva Nastroiiki Semantiko-Orientirovannogo Lingvisticheskogo Protssessora na Vydelenie I Poisk Ob"ektov]. Sbornik IPI RAN, 18 : 119–143.
  12. *Kuznetsov I. P., Kozerenko E. B.* 2008. Linguistic Processor "Semantix" for Knowledge Extraction from Natural Texts in Russia and English. Proceeding of International Conference on Machine Learning, ISAT-2008 : 835–841.
  13. *Kuznetsov I. P., Matskevich A. G.* 2005. English Version of the Automatic Information Detection System for Natural Language Texts [Angloiazychnaia Versiia Sistemy Avtomaticheskogo Vyiavleniia Znachimoi Informatsii iz Tekstov Estestvennogo Iazyka]. Komp'iuternaia Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoi Konferentsii "Dialog 2005" (Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference "Dialog 2005").
  14. *Pasca M., Van Durme B.* 2007. 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) : 2832–2837.
  15. *Punyakank V., D. Roth, W. tau Yih.* 2008. The Importance of Syntactic Parsing and Inference in Semantic Role Labeling. Computational Linguistics, 34(2) : 257–287.