The *problem considered* is to transmit a knowledge represented by natural-language's texts from experts to trainees in the system of computer-aided testing of knowledge.

It's well known that open-form's test's result's interpretation requires the natural language's processing with the respect of synonyms, orthographic mistakes, sense's incompleteness and grammatical incorrectness in trainee's answer. There two-way communication between expert and knowledge base is required for support of actual reflection of reality fragment in the test-developer's language. Also the problem of subject-orientated interpretation of answer be actual here because the results are dependent from the point of view of test's developer's.

The purpose of research is to develop the knowledge structure for synonymy in *natural language (NL)* together with the methods and algorithms for forming and application of this knowledge for the family of tasks:

- sense-similarity's estimation of texts in subject-oriented natural language;;
- computer-aided filling and compression of language's and subject-area's knowledge base;
- seeking a most rational plan for sense's transfer among trainees and experts as different native speakers.

The offered decision is based on the *natural language's usage situation (USNL)* considered as the unit of formalized description of subject area and natural language in common context. Language context fixed by this unit reflects significant objects, relations between them and expression of these relations in texts equivalent-by-sense. The most adequate model of this unit is the *formal context (FC)* known in the formal concept analysis. Here on the basis of formal concept lattice the following classes of semantic relations can be revealed:

- defined by syntactically main word's stem's similarity;
- defined by inflection similarity for main word in syntactic relations what is necessary for their revelation and clustering;
- defined by words's lexical and inflectional compatibility what allows to reveal dependences by analogy with the relations within Russian genitive construction.

Respecting mentioned-above we have the following subtasks to decide a problem of most-rational plan for sense's transfer among trainees and experts:

- to reveal a stem of word as its part being invariant relatively to synonymic transformations;
- to form the criteria for informativeness of words within the frameworks of USNL's context;
- to reveal the set of syntactical links between the words from the phrases defining USNL with the selection of the most projective phrases for USNL's standard's formation.

The decision for the first task is based on the analysis of frequency of occurrence for symbols on different positions in the word concerning its begin and end within the frameworks of USNL. Here the algorithm revealing the stems and inflections for words concerning the USNL was released. The software realization of this algorithm is represented on the Novgorod state university's website. The link

to this release is presented on the personal webpage of author within the frameworks of www.machinelearning.ru. The key feature of algorithm is the grouping of word forms according with the commonality of symbolic prefix. It's necessary to note, that symbols of it has the greatest value of considering frequency relatively to words grouped by common prefix. By analogy with the common prefix the revelation of common suffix was implemented here. This step is important for revelation of reflexive particles.

For revelation of the most compact forms for given sense's expression we entered into consideration a *model of linear structure (MLS) for natural-language phrase*. This model is given on the index set for invariant parts of words with taking into account possible synonyms (according to lemma 5.1). The *sense standard of USNL* is defined by the phrases which are the most projective and having the maximum of the most informative words (respecting synonyms and conversives). *The most informative words* are forming a cluster according to the frequency of occurrence in NL-phrases defining the given USNL. Forming the syntactic links is organized by analogy to learning with a teacher. The first step is revelation of false links on the set of pairs of indexes by means of interview with expert. Each such pair contains indexes lying in the neighborhood in two or more models of linear structures of phrases from defining the USNL. On the basis of initial knowledge the boolean vector is formed for identifying any new link as true or false concerning the links revealed before.

An example of forming the standard is represented on the *slides 12* and *13* for USNL describing the relation between overfitting and empirical risk.

Entering into consideration of sense standards for situations of natural language's usage allows to minimize the volume of information necessary for estimation of affinity of trainee's answer to the correct variant of answer formulated by expert. For example on the *slide 14* are represented the numbers of objects and attributes for the formal contexts representing situations of natural language's usage related to subject area «Mathematical methods of learning by precedents». For comparison on the same slide you can see the representation of such kind for standards of these situations. Thanks to offered idea of USNL considered as the unit of preliminary informational compression is possible to estimate the amount of memory for texts's storing with the respect of possible types of synonymy what is especially actual for knowledge-testing system. Usually for phrase consisting of n words the value $vol(n) = n!$ is taken. Using the standard of USNL here allows to give the upper estimation as $vol_1(n) = l_1 \cdot n$ and lower – as $vol_2(n) = l_2 \cdot n$, where l_1 and l_2 are a numbers of phrases defining USNL and its standard (see *slide 15*).

For numerical estimation of affinity between trainee's answer and correct variant formulated by expert, firstly, and for co-ordination of knowledge formed by different experts on a given subject area, secondly, representation of thesaurus in the form of formal concept lattice is entered into consideration (see *slide 16*). Here the formal concept lattice for a natural language usage's situation be a unit of such thesaurus. In this case the numerical estimation of similarity of natural language usage's situations is determined by the number of attributes be shared by objects of

compared situations concerning the formal context of thesaurus. With the purpose of minimization of data volume necessary for sense's expression at estimating of similarity each situation in thesaurus is represented by its standard. Co-ordinating of data about stems and inflections concerning different situations of natural language's usage on a given subject area allows to make the descriptions for individual situations by objects and attributes more exact, and to increase the exactness of similarity estimations of them (see *slide 18*).

The told can be illustrated by knowledge estimation examples received by knowledge-control system developed by author. Demo-release of this system with source code on Visual Prolog 5.2 presented on the personal webpage of author at www.machinelearning.ru. The *slide 19* presents the interface of system and example of interpretation of trainees answers before the knowledge coordination. The *slide 21* shows the results for the same trainees after the knowledge coordination. Insignificant decrease of estimations of affinity to the correct answer to the *Question No 4* for trainees *Zaitsev E.A.* and *Volkov A.V.* is caused by the replacement of zero inflections revealed earlier for several words from represented in thesaurus.

In released system for each question the USNL for correct answer (with the standard) is put in conformity. For trainee's answer the search of variant closest in symbolic structure is implemented as the first step. Further there is an analysis of words convergence, search of conformities for non-coincident parts of compared phrases within a standard of the correct answer and calculation of estimates of affinity taking into account the found synonyms.

For the purpose of more adaptable interpretation of trainee's answer the numerical estimation of its affinity to the correct variant is calculated for the cases of (see *slide 22*) incomplete answer, orthographic errors (which are admissible) and «excess» words don't appear in any lexico-syntactic links presented in system's knowledge base.

In conclusion let's note, that in current paper all kinds of links between main and dependent word were assumed as equally significant for natural language's phrases similarity's estimation. To apply such estimations in the tasks of testing of professional knowledge relatively to given subject fields it is necessary to re-define the affinity of natural language's usage's situations from viewpoint of fuzzy logic. Here the systems analysis of structure of professional knowledge for the specific area is necessary for the description of membership functions.

Conception of model of linear structure for natural-language phrase can also become more adaptable if probabilities of co-occurrence of words relatively to the texts of given subject area and genre be entered into consideration.