Problems of Ontology Structure and Meaning Optimization and Theirs Solution Methods

Vasyl Lytvyn[1], Oksana Oborska[2], Victoria Vysotska[3], Dmytro Dosyn[4], Andriy Demchuk[5], Yevhen Burov[6], Petro Kravets[7], Nazar Oleksiv

Lviv Polytechnic National University, Lviv, Ukraine

[1]Vasyl.V.Lytvyn@lpnu.ua, [2]Oksana949@gmail.com, [3]Victoria.A.Vysotska@lpnu.ua, [4]Dmytro.Dosyn@gmail.com, [5]Andrii.B.Demchuk@lpnu.ua, [6]Yevhen.V.Brov@lpnu.ua, [7]petro.o.kravets@lpnu.ua4

Abstract. This article are considered optimization problems that occurs during automatic building an ontology about its structure and meaning. A number of this problem have been formed. Has been considered methods and algorithms of their solving.

Keywords: ontology, concept, intelligent agent, structure of ontology, meaning of ontology, problem of optimization.

1 Introduction. Formulation Problem in General

Automatic building an ontology leads to occur of defects in its structure and content and contradiction of its fillings for user information needs [1, 2]. Therefore, such system are needed to be "equip" appropriate set of ontology optimization, namely optimization component. The relevance of the problem of optimization internal representation knowledge bases (KB), which allows to significantly increase the efficiency of intelligent systems, also it mentioned in the Puzankov's work [3]. The following describes the methodology of dynamic ontology construction as well as procedures for its optimization.

The complex structure of relationships between the concepts presented in the ontology, as well as its dynamic content during operation requires the use of certain optimization procedures in order to minimize the response time to requests; not exceeding the space allocated for the ontology; resolving conflicts between data from different sources and meeting other requirements and criteria to be determined. Optimization ontology also performed in order to adapt its content to the information needs of users by removing those items that are rarely used or not used at all, that does not belong to a specific domain. In [4, 5] the expediency of ontology representation in the form of a conceptual graph, the vertices of which are assigned certain semantic and numerical characteristics, is substantiated. It is a weighted graph in which the existence of parallel edges, cycles, loops, duplication of vertices with similar parameters and other features are allowed at the stage of formation. Identification and elimination of such features in order to normalize the graph ontology, as well as optimization of the graph according to certain criteria, taking into account the weight coefficients of edges and vertices should be included in the list of mandatory procedures for ensuring the effective functioning of intelligent decision support systems (IDSS).

In terms of graph theory, structural optimization of an ontology (elimination of conflicts, preservation of integrity, observance of restrictions on the maximum volume) consists in alternation of procedures of addition and reduction of an ontology graph within a given range of values of the number of vertices in the case of maximization of the sum of the importance coefficients of its vertices and edges.

During the functioning of the IDSS there is a constant filling of its ontology with new concepts, which, in turn, requires periodic decision-making on the choice of elements to extract from it (graph reduction) in case of preserving the integrity of its semantic structure. In this regard, when making changes, there is a need to correct the structure and content of the ontology.

For reasons of improving the efficiency of the system as a whole, it is advisable to apply the optimization procedure not every time, but periodically when certain conditions are met (exceeding the time limit or the amount available for filling), alternating the stages of ontology development and the release of redundant elements during optimization. The criteria to be chosen to determine the transition moments from mode to mode should be investigated separately, based on the results of numerical modeling. Consequently, the dynamic content of the ontology and periodic optimization of its structure and content provide restructuring and improvement of the domain model, which is displayed by its ontology.

2 Analysis of Recent Research and Publications

The Gasti project [6] is a system of automatic ontology construction by natural texts. According to this approach, the system starts with a small ontology core containing basic concepts, connections, and operators, and continues to construct the ontology by automatic text processing using symbolic methods, as well as the so-called hybrid approach, consisting of methods of logical, linguistic, pattern management method and semantic analysis.

The peculiarity of the Gasti project is the presence in the system of procedures, the task of which is to periodically clarify and reorganize the ontology. Reorganization consists in clustering (combining similar concepts or splitting the concept into corresponding concepts), transferring common properties up the taxonomy, cutting off unimportant concepts and relationships and excluding some superlatives. The refinement procedure is implemented by four main modules: the "bracketing" module, the "clipping" module, the "merging branches" module and the "discarding redundant paths" module.

Flexibility in relation to user / domain changes in the construction of a dynamic ontology is proposed to provide the assignment of a certain importance to the links and related terms. Importance shows the importance of a weighted element (concept, relationship, or combination thereof) in a given domain from the point of view of the system user. The authors note the need to automate the definition of importance, but today in the system of Gasti it is done manually. The advantages of this model are the adaptability of the ontology to the needs of the user and changes in the subject area, the coverage of a wide range of knowledge, as well as reducing the time and financial costs of building an ontology. The disadvantages include slow query processing and difficulties in learning ontology, due to the specifics of the method of knowledge representation and the nature of learning ontology "from scratch". The absence of a procedure for automatically assigning importance to ontology elements in the system makes it impossible to fully automate it. In General, the review of existing approaches revealed a small number of developed and suitable for practical application methods of dynamic optimization of the ontology of the knowledge base. A new approach to solving this actual scientific and technical task is proposed.

3 Formation of Goals

The aim of the work is to construct optimization problems, the solution of which improves the structure and content of the ontology of the subject area with its automatic restructuring. *The object of research* is the process of optimizing the structure and content of ontology. *The subject of research* is methods and algorithms for solving optimization problems of improving the properties of automatically generated ontology.

4 Selection of Criteria of Optimality of Structure and Content of Ontology of Knowledge Base

We accept such criteria of optimality of structure and content of ontology:

- the physical amount of memory that an ontology occupies;
- performance as the response time of IDSS to external treatment (reaction time to changes in environmental parameters to which the sensitive system);
- completeness of ontology, which can be defined through the average percentage of nontrivial (non-zero) responses to queries to it;
- the integrity of ontology, that is, the absence in its body of mutually contradictory statements and duplication;
- balance of domain expressed as a uniform representation of individual units in the ontology.

Obviously, it is expedient to choose not one criterion, but a certain combination of them to optimize the ontology of BS of applied ISPR, and the choice of the method of such combination has to be carried out empirically, based on real requirements to the system.

COLINS'2020, Volume II: Workshop. Lviv, Ukraine, April 23-24, 2020, ISSN 2523-4013
http://colins.in.ua, online

5 The Integrity of the Ontology Knowledge Base

The ability of an intelligent system to make informed decisions (give answers) on the questions posed to it by developers or users assumes that the system has an ontology that would ensure the validity of such decisions. In particular, the logic of conclusions, the same answer to the same, but differently formulated questions. Such an ontology must satisfy the requirements of integrity.

The concept of integrity combines features or requirements, among which:

- controlled redundancy;
- the connectivity of the graph ontology;
- no conflicting statements.

In knowledge-based systems, redundancy can occur during the completion of their ontology, which consists in the presence of dubious structures: concepts and statements expressed through connections between concepts. When redundant knowledge is required, such systems shall exercise control of redundancy. The connectivity of an ontology graph is a property that means that there is a simple chain between any two vertices of such a graph. Connectivity indicates that all elements of the knowledge base are within the reach of the intellectual system and can be used during the initial response to the appeal to it. When ordering and reducing an ontology, the system must control the preservation of the connectivity condition of its graph and not allow operations that violate this condition.

The graph connectivity test can be carried out using the consequences of the theorem on the evaluation of the number of edges through the number of vertices and the number of connectivity components [7]. If p and q are the number of vertices and edges of the graph, respectively, then the following two conditions must be satisfied:

- if $q>(p-1)(p-2)/2$, then the graph is connected;
- in a connected graph $p-1 \leq q \leq p(p-1)/2$.

When making changes-adding new elements to the ontology, modification, extraction of elements-the system must perform a check on its integrity, that is, the absence of duplicate and/or mutually contradictory statements in it. This procedure can be implemented through a mechanism to identify opposite in meaning reviews test their comparison (comparison) in the case of sequential logical inversion of one of the statements-reviews by the method of resolutions. . In case of coincidence of forward and inverted statements the system receives a signal about violation of integrity and necessity of elimination of contradiction of statements. In the case of conflicting statements, the conflict is resolved by removing the one for which the importance is lower.

In case of exceeding the expected number of reviews in the context, the system generates a procedure for generalizing the concepts that gave such a review, or their properties-opratsovuvachiv in order to reduce the ontology.

The knowledge base will be considered internally consistent if such conditions are met:

- classes, instances, and their attributes have unique names within the scope of certainty;
- all classes in the taxonomy are linked by hierarchical “IS-A " relation-

COLINS’2020, Volume II: Workshop. Lviv, Ukraine, April 23-24, 2020, ISSN 2523-4013
http://colins.in.ua, online

ships, there is only one base class at the highest level of the taxonomic hierarchy;

- all classes contain at least one instance object of a particular class, which provides the functionality of the knowledge view frame model;
- all message handlers of ontology concepts have certain attributes (specific values of formal parameters) at the time of their call.

Some of the features and requirements that the concept of integrity contains (controlled redundancy, graph connectivity, absence of reciprocal statements) are taken into account when formulating restrictions on the ontology structure and optimization problem.

6 Limits on Physical Memory

The system must be implemented on the basis of a certain software and hardware complex, for which there is a real limit on the amount of RAM. On the other hand, the excessive growth of the knowledge base slows down its performance, which can be crucial in the case of systems running in real time.

At the stage of primary formation of the ontology of IDSS at the time of its creation, such a problem does not arise, but during operation, the allotted amount of physical memory is filled, so you have to resort to the procedure of releasing the part of it that is used by the system the least effectively. So, the system works alternately in two modes:

- adding new knowledge to ontology;
- withdrawal from ontology of information which on certain signs represents for the user the least value.

One approach to the selection of features that identify knowledge that is appropriate to withdraw from the ontology is to weigh the concepts and relationships between them during their addition and use during the operation of the system.

To maintain the system in working condition, you need to leave a certain amount of free RAM. The paper selected 10 % of the total. If this benchmark is exceeded during filling, the system enters the optimization mode, during which the sequential selection and extraction of those elements from the ontology for which the ratio to the importance of the space occupied in the RAM will be minimal:

$$J_1 = \min_{1 \le j \le K} \frac{W_j}{m_j}, \tag{1}$$

where W_j is the importance of the element C_j; m_j is the place of the element in RAM; K is the number of ontology concepts.

7 Response Time to External Appeal

The performance determined by the response time of the ISPR to an external appeal

COLINS'2020, Volume II: Workshop. Lviv, Ukraine, April 23-24, 2020, ISSN 2523-4013
http://colins.in.ua, online

can be estimated by the maximum number of arcs of the ontology graph in the possible trajectory of the message propagation between the concepts that are involved in the case of response generation. The performance, determined by the response time of the IDSS to the external appeal, can be estimated by the maximum number of arcs of the ontology graph in the possible trajectory of the message propagation between the concepts that are involved in the case of response generation. The eccentricity of the vertices of the ontology graph (Fig. 1). The eccentricity E_j of a vertex C_j in a connected graph G is the maximum distance from the vertex to the rest of the vertices in graph G. Then the worst performance of the system is:

$$J_2 = \max_{1 \le j \le K} E_j, \quad (2)$$

where E_j is the eccentricity of vertex C_j in a connected graph G, K is the number of ontology concepts.

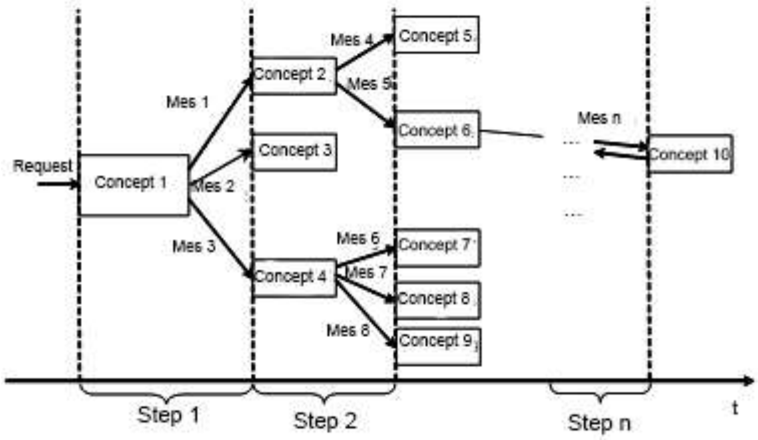

Fig. 1. Scheme of generation of response by concept message handlers to external appeal

8 The Completeness of the Ontology System

Ontology completeness can be defined as the average percentage of nontrivial responses to external requests to the system. By trivial is meant an answer that does not give its recipient new information. In particular, it is possible to rank as trivial the answer of type "information is absent". Another test intelligent system, the completeness of which we know, or a human expert can estimate the percentage of nontrivial answers of a certain system.

The completeness of ontology as the ratio of the completeness of ontology to the number of its concepts is subject to optimization. If the number of compared ontology concepts or knowledge bases are the same, which is the case if they are saturated during operation, it is enough to compare only their completeness.

The principle of determining the completeness of the ontology is based on the methodology of comparison and evaluation of search engines, proposed by the American Institute of standards (NIST) – one of the authoritative bodies of standardization of information technologies in the United States. The methodology uses a corpus of test questions and documents accumulated during the text search systems assessment conferences (TREC and text retrieval evaluation conference) held by NIST.

Such criterion is integral and does not allow to carry out regular optimization of structure of ontology on its basis, and is intended for an assessment and comparison of functioning of information systems as a whole.

9 Balance Subject Area

During automatic ontology rearrangement, it is possible that as a result, the granularity of domain concepts, represented by the number of classes, subclasses, their instances and properties, for one class may differ significantly from the display of another class. Balanced domain is expressed in a uniform representation of individual sections in the ontology. The requirement of balance may be in front of commercial universal IDSS, the scope of which cannot be determined a priori. The formal criterion of the balanced representation of the concept-class in the ontology of IDSS can be the variance of the importance of its subclasses:

$$\sigma_j^{<k>} = \overline{\left| \overline{W_i^{<k+1>}} - W_i^{<k+1>} \right|^2}, \ i = \overline{1..n}, \ j = \overline{1..K}, \tag{3}$$

where $W_i^{<k+1>}$, $\overline{W_i^{<k+1>}}$ there are importance and averaged importance of K + 1-level subclasses, respectively; n is the number of subclasses in the J-th class, N is the number of classes in the ontology.

When optimizing the ontology, the balance criterion can be used when choosing the subject of exemplary texts to complement the ontology. In this paper, the ontology content optimization algorithm does not contain such a criterion of the objective function, so it is not considered in detail.

10 Formulation and Solution of Optimization Problem

We formulate an optimization problem, which will be the basis of the ontology completion and optimization procedures. The automatic generation of the ontology determines the need to carry out its local optimization during the filling, and the ball-at the ordering stage, when the filling of the prize is completed before the completion of the optimization procedure. The ontology optimization method contains the problem of normalization of its structure and the problem of content optimization (Fig. 2), in addition:

1) the elimination of parallel edges, duplication of vertices with the same parame-

COLINS'2020, Volume II: Workshop. Lviv, Ukraine, April 23-24, 2020, ISSN 2523-4013
http://colins.in.ua, online

ters and other features of the structure of the graph of ontology, which can violate its integrity and reduce the efficiency of the functioning of the intellectual system, is the task of normalizing the structure of ontology;

2) optimization of the content part of the ontology is performed in order to increase its performance and information saturation due to restrictions on the physical memory of the system.

It is assumed that the solution of these problems must be divided in time, and in order to preserve the integrity of the ontology, one must first perform its structural check, and only then - optimize the content part by sequential reduction of its graph to fulfill the requirements of the selected criteria by maximizing the sum the importance of the vertices and edges of such a graph.

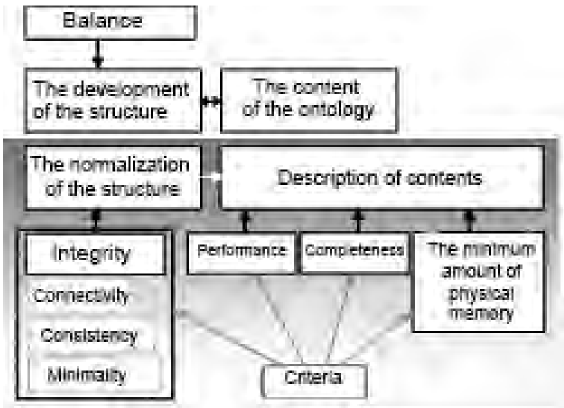

Fig. 2. Application of ontology optimality criteria at the stages of its filling and ordering

Some tasks are shown in the figure, namely the development of the structure and content of the ontology (on a white background) do not require optimization, so it is not the subject of detailed consideration in this work.

10.1 The Task of Normalization of the Structure

The above criteria are defined to ensure the integrity of the ontology-coherence, consistency (absence of contradictory statements) and minimality (controlled redundancy), which are the basis of the optimization problem. One of these criteria – connectivity - will act as a constraint that must be satisfied, but the minimality and consistency criteria define two separate optimization problems. The problem of minimizing the structure of the ontology graph is based on a typical optimization problem of graph theory about finding a minimum skeleton, which consists in finding a skeleton of minimum importance in a weighted graph. But the problem of ensuring consistency in the structure of the graph is effectively solved by applying the method of resolutions. The order of application of such optimization procedures is important, since the structure minimization is partially achieved during the elimination of logical inconsistencies. So, consider them sequentially in order of application.

10.2 Ensuring Consistency of Ontology Structure by Resolution Method

The presence of negative statements in the structure of ontology leads to internal logical conflicts, violating its integrity, so the system must be able to timely identify and remove those that are recognized as erroneous. Such a problem can be effectively solved by applying the method of resolutions – the classical method of automatic proof of theorems. This method is based on proving from the opposite to the consistency of a set of statements. For this formula in the predicate calculus using the corresponding transformations are reduced to disjunctive form, that is presented as a set of sentences of a certain class of well-formed formulas, unquantifiable disjunction of literals, each of which is either a predicate or the negation of a predicate.

The disadvantage of the resolution method in its direct application can be attributed to the "blind" search, since the resolution method is a iterative procedure. Its direct application is inefficient, since for large ontologies it generates a significant number of unpromising resolutions. To prevent "exponential explosion" it is necessary to apply more effective modifications of the method: semantic resolution, loc-resolution, linear resolution. The essence of these modifications is the introduction of certain criteria by which disjuncts participating in the next resolution are selected. These modifications in General do not eliminate the backtracking search, but allow to significantly reduce this search and make the method more suitable for practical use. The linear resolution algorithm is applied in the paper, taking into account its efficiency and ease of implementation [8].

10.3 Normalizing the Structure by Searching for the Minimum Skeletal

Normalization of the ontology structure, first of all, is to ensure its minimality, that is, the absence of information redundancy. In the case of representing the ontology structure by a weighted graph, where the importance of the edge reflects the importance of the relationship reflected by it, and redundancy can manifest itself in the form of parallel edges, the procedure for eliminating such features may consist in sequentially removing the edges with minimal importance while maintaining the connectivity of the entire graph. To solve this problem, it is proposed to use an algorithm for constructing the minimum frame of the graph [7]. A skeletal is a subgraph of some graph $G(V, E)$ containing all its vertices and being a tree. When the edges of a graph have importance, one can construct a minimal skeletal tree (MST – Minimal Spanning Tree) of a weighted graph – it is a skeletal tree whose importance (the sum of the importance of its edges) does not exceed the importance of any other skeletal tree.

In our case, instead of allocating edges with minimal importance, it is necessary to allocate edges with maximum importance, taking into account the peculiarities of providing a semantic network with different types of semantic connections by a conceptual graph.

A common graph of a knowledge base ontology is like a graph of a semantic network consisting of subgraphs that share vertices-entities, but different types of edges that reflect different semantic relationships between these entities. The procedure for constructing a minimal skeletal tree is performed separately, sequentially for each of the subgraphs allocated by the type of semantic relations.

Therefore, the overall strategy of constructing a minimal skeleton in the task of minimizing the structure of the graph is as follows: at each step, an edge is added to the fragment of the skeletal tree formed in the previous step, the highest in importance among those edges that already connect the construct. This is a fragment with vertices that have not yet been added to this fragment. In order to economically implement the steps of this process, we associate with each vertex $x \in VG$ the two labels β (x) and γ (x) are the importance of the most important edge connecting the vertex x to the constructed fragment of the maximum skeleton, and the name of the second vertex of this edge, respectively. The β(x) and γ(x) labels allow you to quickly find the edge of maximum importance at each step.

Let us assume that the graph G, which reflects the structure of the ontology of the knowledge base of a certain semantic type, is given by the importance matrix. Then the algorithm for constructing the minimum skeleton will have the form described below.

1. We select any vertex of the graph of the knowledge base G and find the edge (connection) adjacent to it with the greatest importance. At the same time, let $ET := \varnothing$, $VT := \{a\}$, where VT, ET are the sets of vertices and edges of the minimal backbone fragment being constructed; a is any vertex from VG.
2. Assign labels to the remaining vertices $x \neq a$ on the principle $\beta(x)=W(x, a)$, if $x \in N_a$, and $\beta(x)=0$, if $x \notin N_a$, $\gamma(x)=a$, N_a is the environment of vertex a.
3. Choose a vertex $v^* \in VG \backslash VT$, the closest to the fragment by condition: $\beta(v^*) = \min_{v \in VG/VT} \beta(v)$.
4. Increase the fragment. Among the unused vertices we find the vertex with the maximum value of x and connect it to the corresponding vertex of the skeleton. To the rest of the unused vertices, re-assign the value β (x) from γ(x) taking into account the increase of the backbone, that is, if β(v)>W(v*,v), then β(v):=W(v*,v), γ(v):=v*. If β (v)≤W (v*, v), then the vertex label v is not changed. Proceed to point 3.
5. Repeat the operations until there are no unoccupied vertices, that is, if $|VT|=n$, the procedure must be completed. The edges contained in the plural ET are the minimum skeleton.

On Fiig. 3 shows the graph G and the sequence of construction of fragments of the minimum skeleton, which are formed after each iteration of the algorithm. Since each time an edge is added to ET, one end of which belongs to VT and the other does not, then the graph T=(VT, ET) remains a tree at each step. After the algorithm completes, this tree will be the backbone, since the algorithm stops working when VT=VG.

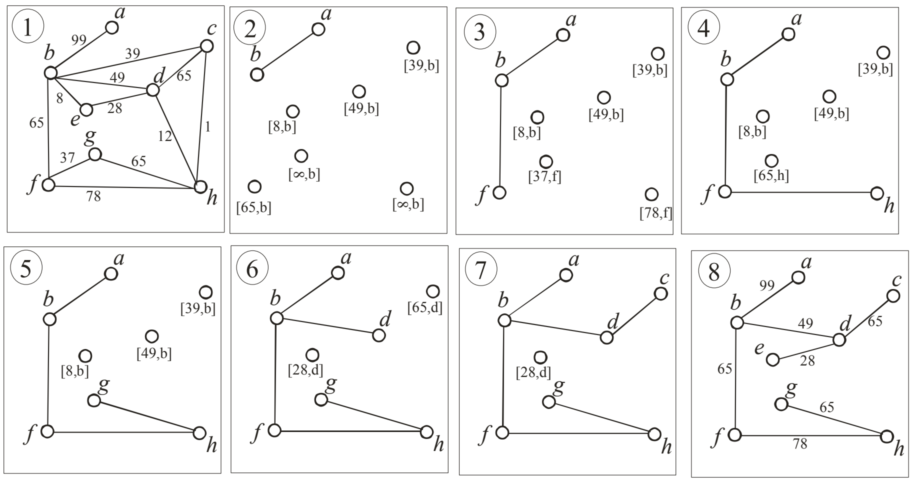

Fig. 3. The sequence of construction of the minimal skeleton of the graph G

We will show that this algorithm builds a minimal spanning tree for the time $O(|G|^2)$. The one-time execution of paragraph 1 requires O (|G|) time. The same amount of time is needed to update the labels in step 4, and the fragment increase is performed in time O (1). Since each of the paragraphs 2-5 is executed n-1 times, the estimation of the complexity of the algorithm is $O(|G|^2)$.

10.4 The Problem of Optimization of Content

The criteria by which the optimal content of the ontology is achieved are selected and justified above: performance, completeness and minimum amount of physical memory. The procedure for optimizing ontology is to consistently remove from it the concepts that are of least importance. At this stage, based on these criteria, it is necessary to formulate an optimization problem, that is, to formulate a target function and impose the appropriate restrictions arising from the method of implementation of IDSS. Since the optimal content of the ontology of such a system is estimated by contradictory indicators, it is not possible to "bring them to a common denominator" at the same time. In this case, we have a situation of uncertainty of goals, which can be eliminated by using one of the methods of eliminating uncertainty (benchmarks, the method of linear convolution, the introduction of metrics in the space of objective functions, Pareto compromise).

So, first you need to solve the problem of finding the optimal number of concepts in the ontology, and then, using the results obtained, to implement the procedure of periodic alternately adding new concepts to the ontology and extracting elements from the ontology, the semantic value of which is less for the system. This procedure should be the basis of the method of adaptation to the ontology of the user-defined domain, as a functional element of the corresponding intellectual system. The problem of extracting the least important elements from the ontology can be formulated as a discrete linear optimization problem, as evidenced by the results of simulation of the complement and weighting procedures. A similar condition of such an optimization problem is the problem of the solution [7]. Only here the share of ontology concepts subject to withdrawal will act as a backpack.

Let's consider the limitations that are imposed on such an optimization problem, determine the optimal performance / completeness volume of the ontology and formalize, giving in the form of an algorithm, the solution of the problem of selecting elements to extract from the ontology.

10.5 The Imposition of Constraints on the Optimization Problem

As a limit on the number of concepts of the optimized ontology, we introduce the parameter of the minimum allowable amount of free RAM. The choice of this parameter is due to the need to maintain the system in working condition while supplementing its ontology. Its value depends on the specific conditions of implementation of the intelligent system (hardware and software). For certainty, the value of the parameter will take 1 % of the total memory occupied by the ontology:

$$\Delta M = 0,01\sum_{j=1}^{K} m_j \quad (4)$$

where m_j is the volume that occupies the j-th element in the IDSS RAM.

If this benchmark is exceeded during filling, the content optimization process starts, during which the elements selected in the above manner are sequentially removed from the ontology.

The optimization process can be carried out both in case the target function reaches the optimal value and in case the optimization parameters (the target function) reach critical, that is, restrictive values. In the case of optimizing the content of the ontology, the removal of concepts that carry a small semantic load, and therefore have the least importance among all the concepts that are considered, stops when the relative amount of free RAM exceeds the second threshold value-10 % . The choice of a numeric value also depends on the conditions of the IDSS implementation and, in particular, its ontology.

10.6 The Problem of Finding the Optimal Number of Ontology Concepts

During the operation of ontology there is a dilemma of ensuring, on the one hand, the most complete reproduction of the semantic structure of the analyzed texts in the case of their representation in the form of conceptual graphs, and on the other, the maximum performance of the system of their intellectual analysis.

A well-known lexical-semantic source suitable for building a General-purpose ontology is the WordNet database, developed and maintained by scientists at Stanford University (USA). It is de facto considered the standard of semantics of the English language and allows you to reproduce the" IS-A " – a hierarchy of concepts with a volume of about 100 thousand words. However, the complexity of the algorithms that have to be implemented on ontology graphs in order to recalculate the importance coefficients of its elements and construct semantic images of translated texts, as shown by the research, is power, in particular, the search for the semantic center of the importance of the conceptual graph is $O(n^3)$. Therefore, the dimension of the ontology significantly affects the performance of IDSS.

There is a class of application tasks that require real-time processing of text data (for example, monitoring and filtering continuous mail traffic). At each subsequent moment a new batch of data arrives in need of analysis. In such cases, there is a dilemma between choosing the completeness of such analysis and performance. Reducing the dimension of the ontology to certain optimal boundaries is an important optimization task of the stage of its development, at which it is possible to take into account the peculiarities of the operating conditions of the system, giving preference to either the speed or accuracy of the resulting IDSS solution. The accuracy of the resulting IDSS solution depends on the completeness of the representation in the ontology of the defined SOFTWARE. In the framework of this work, a simple linear model of dependence is taken as a basis, which provides us with an upper bound for estimating the accuracy drop caused by a decrease in the volume of the ontology:

$$\tau = \frac{r_e}{r_m} = f(G) \approx \xi G, \quad (5)$$

where $r_e = |R_e|$ is the power of the set R_e of right decisions issued by the IDSS (for example, according to an expert assessment of a particular domain) found by the appropriate method; $r_m = |R_m|$ is the power of the set R of all system-issued solutions found by the corresponding method; $R_e \subseteq R_m$; G is the relative number of concepts in ontology; G = K/N (K – number of ontology concepts; N - number of domain concepts); ξ is the linear coupling coefficient.

Therefore, the balance between speed and accuracy of the search can be reduced in the model approximation to the search for the optimal ratio of speed and completeness. The practical implementation of the ontology as a semantic network of frames, the tops of which are placed the object modules and links implemented on the basis of opracowac of these modules determines the need to evaluate the system performance as the system response time to external appeal. An arbitrary number of object modules can participate in the processing of such an appeal through an appeal to each other. The reaction time is generally determined by the slowest component of the system. Therefore, the performance of such a system can be estimated because of the averaged slow reaction of its components.

So, as already mentioned, to solve such a problem, it is necessary to determine the optimal number of elements in the ontology of the IDSS knowledge base on the basis of performance and completeness criteria. For this purpose we write the target function

$$f = \bar{E} + \frac{1}{G} \to \min . \tag{6}$$

The eccentricity of the vertex of the ontology graph, whose connectivity is ensured by the mandatory existence of "IS-A" - connection between all elements of the ontology, will be no more than the sum of the levels of the taxonomic hierarchy between the given and the root elements and the maximum number of levels in the ontology. An arbitrary additional semantic relationship between elements of an ontology reduces the average eccentricity of the vertices of its ontology. So, for a certain statistical distribution of the number of elements on the levels of the "IS-A" hierarchy, the average eccentricity of the vertices of the graph, reflecting the ontology:

$$\bar{E} = \frac{N\sum_{i=1}^{k}\left[P(i)\cdot\left(i+k\right)\right]}{K}, \; K = N\sum_{i=1}^{k}P(i), \tag{7}$$

where k is the number of levels in the graph; K is the number of concepts in the ontology; N is the number of concepts in the domain dictionary; $P(i)$ is the probability of finding an ontology element at the i-th level of the taxonomic hierarchy of concepts, established on the basis of the analysis of the lexical database WordNet.

It is established that for such an actual distribution of the specified optimality criteria and the dictionary, which contains 100,000 concepts, the minimum of the target function corresponds to the optimal number of 31,000 concepts in the adaptive ontology of IDSS.

10.7 Representation of the Problem of Extracting Elements from Ontology as the Inverse of the Scapular Problem

For practical application of discrete optimization problem of ontology instead of integral criteria of optimality of ontology as a whole it is necessary to apply the criteria characterizing each separate element of ontology and give the chance to make on it the decision on necessity or inexpediency of its introduction to ontology (an exception from it). For this purpose, criteria have been developed and adapted for application in this method.

Let assume that each element of ontology takes the same amount of memory and therefore the total volume occupied by linearly dependent on the amount specified in the ontology concepts . The mean eccentricity of the vertices of an ontology graph is taken into account in the self-importance of classes and the importance of their subclasses, falling with the level. If we take into account the fact that the ontology contains more than 10-100 thousand individual concepts located in the taxonomic 15-25-level hierarchy, the relationship between the importance of concepts close to the root and the importance of concepts of the lower levels is several orders of magnitude. Thus, for semantic problems, the removal or introduction of the ontology of lower-level concepts during optimization does not significantly change the importance of other concepts. Therefore, we can use the model assumption of the independence of the average importance of ontology concepts from their total number, and the linear discrete problem of content optimization can be represented as the problem of the scapular. Backpack will act and the proportion of the volume of ontology, which should include the least valuable elements in order to further their withdrawal.

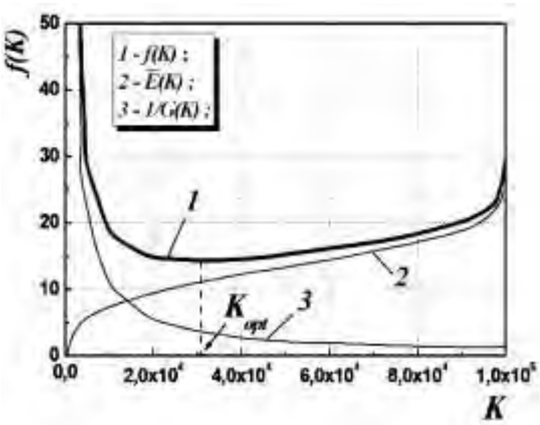

Fig. 4. The objective function of finding the optimal number of ontology concepts

Applying a simple search for an ontology with *n* elements, the upper estimate of the number of possible combinations will be *2n*. in addition, to solve such a problem, it is enough to apply dynamic programming methods, a greedy algorithm or the method of branches and boundaries.

During the study of the effectiveness of the implementation of individual methods, the choice was stopped on the use of a greedy algorithm. This algorithm has complexity *O(n log (n))* and guarantees finding of the solution not worse than twice from optimum that is acceptable for this problem.

Let for an ontology with n elements the excess volume is *N=1 / 10M*, where *M* is the maximum allowable volume. If the introduction of the *i*-th element in the" duffel bag "" x_i, $0 \le x_i \le 1$ determines the benefit x_i/W_i, then the optimal filling of the "duffel bag" is considered to be such a filling with elements that maximizes the total total benefit. The problem can be formulated as follows: maximize $\sum_{i=1}^{n} \frac{1}{W_i} x_i$ such and elements for which:

$$\sum_{i=1}^{n} m_i x_i \le N \, , \; W_i > 0 \, , \; m_i > 0 \, , \; i = 1, \ldots n \, , \; \begin{cases} x_i = 0, \text{ if remains,} \\ x_i = 1, \text{ if withdrawn.} \end{cases} \quad (8)$$

The task is simplified if we accept,, which usually corresponds to the conditions of the ontology implementation.

The implementation of the greedy algorithm in this problem is reduced to the sequential execution of the following operations (Fig. 5):

1) search for the item with the least importance.;
2) adding it to the set of those that are withdrawn (assignment xi=1);
3) calculate and check the conditions (8);
4) if condition (8) is met, proceed to clause 1;
5) in case of violation of the terms, stop the search and delete the selected items.

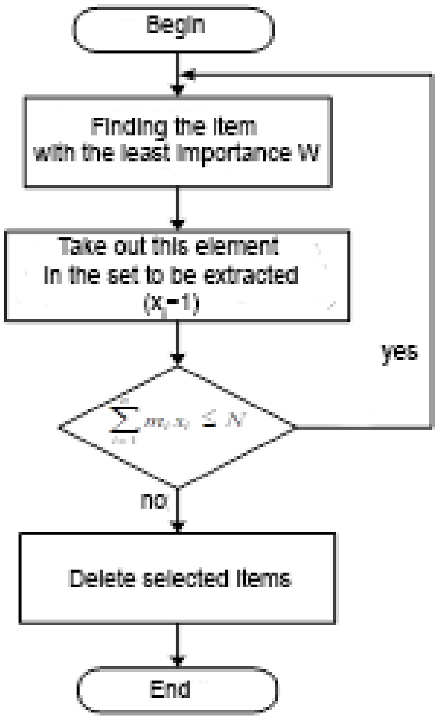

Fig. 5. Block diagram of greedy algorithm implementation

The time estimate of the algorithm depends entirely on the time estimate of the ordering algorithm used, since it takes *O(n)* time to implement the filling strategy itself after the items are ordered.

11 Conclusions

Automatic filling of ontology causes the urgency of the problem of its periodic optimization. The set of criteria of optimality of structure and content of ontology which reflect requirements to operational characteristics and technical possibilities of realization of the corresponding systems is proved. The task of ontology optimization is proposed to be divided into the task of structure normalization and the task of content optimization. Elimination of parallel edges, loops, loops, duplication of vertices with similar parameters and other features of the structure of the graph of ontology, namely: logical contradictions that violate its integrity and reduce the efficiency of functioning, is the task of normalization of the structure. During the solution of this problem the method of calculation and algorithm of construction of the minimum skeleton is applied. Optimization of the ontology content is performed to increase its information saturation and ensure adaptation to the information needs of the user through periodic reduction of its volume to the specified limits by removing the elements whose semantic value is the least. It is shown that the optimization problem can be reduced to a discrete optimization problem on the shoulder, the algorithms for the effective solution of which are known.

References

1. Lytvyn, V., Vysotska, V., Dosyn, D., Burov, Y.: Method for ontology content and structure optimization, provided by a weighted conceptual graph. In: Webology, 15(2), 66-85. (2018)
2. Lytvyn, V., Vysotska, V., Pukach, P., Vovk, M., Ugryn, D.: Method of functioning of intelligent agents, designed to solve action planning problems based on ontological approach. In: Eastern-European Journal of Enterprise Technologies, 3/2(87), 11-17. (2017)
3. Lytvyn, V., Vysotska, V., Dosyn, D., Lozynska, O., Oborska, O.: Methods of Building Intelligent Decision Support Systems Based on Adaptive Ontology. In: Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP, 145-150. (2018).
4. Lytvyn, V., Vysotska, V., Veres, O., Rishnyak, I., Rishnyak, H.: Classification methods of text documents using ontology based approach. In: Advances in Intelligent Systems and Computing, 512, 229-240. (2017)
5. Vysotska, V., Lytvyn, V., Burov, Y., Berezin, P., Emmerich, M., Fernandes, V. B.: Development of Information System for Textual Content Categorizing Based on Ontology. In: CEUR Workshop Proceedings, Vol-2362, 53-70. (2019)
6. Lytvyn, V., Vysotska, V., Burov, Y., Demchuk, A.: Architectural ontology designed for intellectual analysis of e-tourism resources. In: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 335-338. (2018)
7. Lytvyn, V., Vysotska, V., Rusyn, B., Pohreliuk, L., Berezin, P., Naum O.: Textual Content Categorizing Technology Development Based on Ontology. In: CEUR Workshop Proceedings, Vol-2386, 234-254. (2019)
8. Kravets, P., Burov, Y., Lytvyn, V., Vysotska, V.: Gaming method of ontology clusterization. In: Webology, 16(1), 55-76. (2019)
9. Shu, C., Dosyn, D., Lytvyn, V., Vysotska V., Sachenko, A., Jun, S.: Building of the Predicate Recognition System for the NLP Ontology Learning Module. In: International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS, 2, 802-808. (2019)
10. Malyeyeva, O., Nosova, N., Fedorovych, O., Kosenko, V.: The Semantic Network Creation for an Innovative Project Scope as a Part of Project Knowledge Ontology. In: Computational linguistics and intelligent systems, COLINS, 301-311. (2019)
11. Burov, Y., Vysotska, V., Kravets, P.: Ontological Approach to Plot Analysis and Modeling. In: Computational linguistics and intelligent systems, COLINS, 22-31. (2019)

12. Lytvyn, V., Oborska, O., Vysotska, V., Dosyn, D., Demchuk, A.: Ontology Using for Decision Making in a Competitive Environment. In: Computational Linguistics and Intelligent Systems, COLINS, 2, 17-27. (2018)
13. Fernandes, V.B.: Ontologies Quality Assessment and Optimization. In: Computational Linguistics and Intelligent Systems, COLINS, http://colins.in.ua/wp-content/uploads/2018/07/OntologiesQualityAssessmentandOptimization.pdf, 1-11. (2018)
14. Orobinska, O., Chauchat, J.-H., Sharonova, N.: Methods and models of automatic ontology construction for specialized domains (case of the Radiation Security). In: 1st International Conference Computational Linguistics and Intelligent Systems, COLINS, 95–99. (2017)
15. Lytvyn V., Vysotska V., Wojcik W., Dosyn D.: A method of construction of automated basic ontology. In: 1st International Conference Computational Linguistics and Intelligent Systems, COLINS, 75–83. (2017)
16. Kuprianov, Ye.: Semantic state superpositions and their treatment in virtual lexicographic laboratory for spanish language dictionary. In: 1st International Conference Computational Linguistics and Intelligent Systems, COLINS, P. 37–46. (2017)
17. Литвин В.В. Розроблення методів та засобів побудови інтелектуальних систем опрацювання інформаційних ресурсів з використанням онтологічного підходу / В.В. Литвин, В.А. Висоцька, Д.Г. Досин, М.Г. Гірняк // Інформаційні системи та мережі. Вісник Національного університету "Львівська політехніка". – № 832. – Львів, 2015. – Стор. 295-314.
18. Литвин В.В. Особливості рубрикації текстових документів з використанням онтології / В.В. Литвин, В.А. Висоцька, Р.М. Оливко, Т.М. Черна // Міжнародна наукова конференція "Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту (ISDMIT'2016)", Залізний Порт, Україна. – 25-28 травня 2016. – Стор.292-295.
19. Литвин В.В. Методи та засоби опрацювання інформаційних ресурсів на основі онтологій: монографія / В.В. Литвин, В.А. Висоцька, Д.Г. Досин. – Львів: ЛА «Піраміда», 2016. – 404 с.
20. Литвин В.В. Побудова онтології архітектурних термінів / В.В. Литвин, О.І. Ремешило-Рибчинська, В.А. Висоцька // Відбір і обробка інформації. – 44 (120). – Львів 2017. – С. 90-96. – ISSN 0474-8662.
21. Досин Д.Г., Висоцька В.А., Литвин В.В. Побудова системи підтримки прийняття рішень на базі адаптивної онтології // Обчислювальні методи і системи перетворення інформації: зб. пр. V-ї наук.-техн. конф., (Львів, 4-5 жовтня 2018 р.) - Львів: ФМІ НАН України, 2018. - Вип. V. - С. 135-138
22. Висоцька В. А. Методи та засоби функціонування систем підтримки прийняття рішень на основі онтологій: монографія / В.А. Висоцька, Д.Г. Досин, Х.І. Микіч, І.І. Завущак, З.Л. Рибчак. – Львів: Видавництво «Новий світ – 2000», 2019. – 334 с.
23. Пасічник В.В. Математична лінгвістика. Книга 1. Квантитативна лінгвістика / В.А. Висоцька, В.В. Пасічник, Ю.М. Щербина, Т.В. Шестакевич // Навчальний посібник з грифом Міністерства освіти і науки, молоді та спорту України. – Львів: видавництво «Новий світ-2000», 2012. – 359 с.
24. Пасічник В. В. Математична лінгвістика. Книга 2. Комбінаторна лінгвістика: навчальний посібник / В. В. Пасічник, Ю. М. Щербина, В. А. Висоцька, Т. В. Шестакевич. – Львів: Видавництво Львівської політехніки, 2019. – 250 с.

25. Берко А.Ю. Лінгвістичний аналіз текстового комерційного контенту / А.Ю. Берко, В.А. Висоцька, Л.В. Чирун // Інформаційні системи та мережі. Вісник Національного університету "Львівська політехніка", № 814.- Львів 2015 – Стор. 203-227.
26. Литвин В.В. Метод визначення семантичної метрики на основі тезаурусу предметної області / В.В. Литвин, В.А. Висоцька, Р.М. Оливко // V Всеукраїнська науково-практична конференція «Інтелектуальні системи та прикладна лінгвістика». – Харків, 2016. – 14 квітня 2016 р. – Національний університет «Харківський політехнічний інститут». – Стор. 10-12.
27. Бісікало О.В. Метод лінгвістичного аналізу україномовного комерційного контенту / О.В. Бісікало, В.А. Висоцька // Інформаційні системи та мережі. Вісник Національного університету "Львівська політехніка", № 854.- Львів 2016 – Стор. 185-204.
28. Vysotska V. Analytical methods for commercial web content processing of information resource in electronic business systems / Victoria Vysotska, Lyubomyr Chyrun // MEST Journal (Management Education Science & Society Technologie). – Vol.2 No.2. – PP. 285-300 [Online]. – ISSN 2334-7171, ISSN 2334-7058 (Online), DOI 10.12709/issn.2334-7058. This issue: DOI 10.12709/mest.02.02.02.0. – Режим доступу: http://mest.meste.org/MEST_2_2014/4_29.pdf.
29. Vysotska V. Set-theoretic models and unified methods of information resources processing in e-business systems / Victoria Vysotska, Lyubomyr Chyrun // Applied Computer Science. ACS journal. – Volume 10, Number 3. – Poland, 2014. – ISSN 2353-6977 (Online), ISSN 1895-3735 (Print). – PP. 5-22 [Online]. – Режим доступу: www.acs.pollub.pl, www.acs.pollub.pl, http://www.acs.pollub.pl/pdf/v10n3/1.pdf.
30. Vysotska V. Linguistic analysis and modelling semantics of textual content for digest formation / Victoria Vysotska, Lyubomyr Chyrun // MEST Journal (Management Education Science & Society Technologie). – Vol.3 No.1. – PP. 127-148 [Online]. – ISSN 2334-7171, ISSN 2334-7058 (Online), DOI 10.12709/issn.2334-7058. This issue: DOI 10.12709/mest.02.02.02.0. – Режим доступу: http://mest.meste.org/MEST_1_2015/5_15.pdf.
31. Vysotska V. Features of Text Categorization of Commercial Content / Victoria Vysotska // Computer Science and Information Technologies: Proc. of the IX-th Int. Conf. CSIT'2014, 18-22 November, 2014, Lviv, Ukraine..– Lviv: Publishing Lviv Polytechnic, 2014.– P.5-8.
32. Berko A. Linguistic Analysis for the Textual Commercial Content / Andriy Berko, Victoria Vysotska, Lyubomyr Chyrun // Computer Science and Information Technologies: Proc. of the IX-th Int. Conf. CSIT'2014, 18-22 November, 2014, Lviv, Ukraine..– Lviv: Publishing Lviv Polytechnic, 2014.– P.11-14.
33. Shestakevych T. Modelling of semantics of natural language sentences using generative grammars / Tetiana Shestakevych, Victoria Vysotska, Lyubomyr Chyrun, Liliya Chyrun // Computer Science and Information Technologies: Proc. of the IX-th Int. Conf. CSIT'2014.– Lviv: Publishing Lviv Polytechnic, 2014.– P.19-22.
34. Andrunyk V. The Peculiarities of Electronic Digest Formation / Vasyl Andrunyk, Victoria Vysotska, Lyubomyr Chyrun // Computer Science and Information Technologies: Proc. of the IX-th Int. Conf. CSIT'2014, 18-22 November, 2014, Lviv, Ukraine.– Lviv: Publishing Lviv Polytechnic, 2014.– P.25-28.

    Kis I. Features of the Content-Analysis Method in Processing Online Newspaper Articles / Iaroslav Kis, Victoria Vysotska, Liliya Chyrun, Vasyl Foltovych // Computer Science and Information Technologies: Proc. of the IX-th Int. Conf. CSIT'2014, 18-22 November, 2014, Lviv, Ukraine..– Lviv: Publishing Lviv Polytechnic, 2014.– P.39-42.

35. Берко, А.Ю. Intranet архітектура інтелектуальних систем електронного навчання / А.Ю. Берко, В.А. Висоцька // Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”. – Львів 2001. - № 438. – Стор.3-10.
36. Інтерактивна взаємодія та зворотній зв’язок в системі дистанційного навчання / Р.О. Голощук, В.А. Висоцька // Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”. – Львів 2002. – № 464. – Стор.44-53.
37. Висоцька, В.А. Система опрацювання структури електронного підручника / В.А. Висоцька // Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”. – Львів 2003. – № 489. – Стор.49-63.
38. Голощук, Р.О. Математичне моделювання процесів дистанційного навчання / Р.О. Голощук, В.В. Литвин, Л.В. Чирун, В.А. Висоцька // Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”. – Львів 2003. – № 489. – Стор.100-109.
39. Берко, А.Ю. Алгоритми опрацювання інформаційних ресурсів в системах електронної комерції / А.Ю. Берко, В.А. Висоцька, Л.В. Чирун // Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”. – Львів 2004. – № 519. – Стор.10-20.
40. Берко, А.Ю. Проектування навігаційного графу Web-сторінок бази даних систем електронної комерції / А.Ю. Берко, В.А. Висоцька // Комп’ютерні науки та інформаційні технології. Вісник Національного університету “Львівська політехніка”. – Львів 2004. – № 521. – Стор.48-57.
41. Берко, А.Ю. Застосування OLAP – технології в системах електронного бізнесу / А.Ю. Берко, В.А. Висоцька // Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”. – Львів 2005. – № 549. – Стор.3-15.
42. Берко, А.Ю. Застосування технологій електронного бізнесу у видавничій справі / А.Ю. Берко, В.А. Висоцька, А.А. Висоцький // Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”. – Львів 2007. – № 589. – Стор.14-26.
43. Висоцька, В.А. Аналіз та оцінювання ступеня ризиків в електронній комерції / В.А. Висоцька, І.В. Рішняк, Л.В. Чирун // Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”. – Львів 2007. – № 589. – Стор.56-68.
44. Берко, А.Ю. Аналіз і класифікація моделей систем електронної комерції / А.Ю. Берко, В.А. Висоцька // Автоматизовані системи управління та прилади автоматики. Науково-технічний журнал. – Харків 2006. – № 137. – Стор.48-55.
45. Берко, А.Ю. Моделі та методи проектування інформаційних систем електронної комерції / А.Ю. Берко, В.А. Висоцька // Автоматизовані системи управління та прилади автоматики. Науково-технічний журнал. – Харків 2007. – № 138. – Стор.55-66.
46. Берко, А.Ю. Застосування мереж Петрі для моделювання систем електронної комерції / А.Ю. Берко, В.А. Висоцька // Радіоелектроніка. Інформатика. Управління. Науково-технічний журнал. – Запоріжжя 2007. – № 2. – Стор.26-33.

Висоцька, В.А. Моделі функціонування систем електронної комерції: тези науково-практичної конференції, 12 квітня 2007 р. / В.А. Висоцька // 11-й Міжнародний молодіжний форум “Радіоелектроніка та молодь в XXI столітті”. Харківський національний університет радіоелектроніки. – Харків 2007. – Стор.194.