

AN UNIVERSAL SYSTEM FOR CREATING AND APPLYING. ELECTRONIC ENCYCLOPEDIAS FOR THE INTERNET

Kostenko K.I. Levitskij B.E.
Kuban state university
Krasnodar
(8612)699550/(8612)699550/ kostenko@kubsu.ru

Electronic encyclopedias are intended both for the general user and the professionals seeking consultation. Those encyclopedias has a special place in the electronic libraries since it has the possibility to construct search systems that provides automatic answers to intellectual inquiries of the information consumers.

The special approach to the organization of document storage as knowledge bases, makes the application of electronic encyclopedias for the users problems decision possible, with the help of knowledge processing procedures.

The work on creating the universal system VEDA to construct electronic document storehouses and mechanisms of their processing are conducted in the Internet center of the Kuban University.

General principles of VEDA.

The following principles of the storehouse organization for documentary resources are realized in the VEDA project [1]:

- hierarchical representation of the structure of the documents placed in the system;
- an opportunity of creation and application of the documents containing fields with professional knowledge;
- realization of intellectual search mechanisms based on the concept of associations;
- separation of the documents internal content and the forms of its visualization for the users;
- organisation of an information upkeep for long-term interests of the users.

Electronic encyclopedias developed in the VEDA system consist of structured articles, which are created in view of features of knowledge area and accepted forms of the documents.

Authorship protection and an opportunity of document accommodation for articles with different points of view, or fragments of descriptions of the same objects are supported in the electronic encyclopedia, created in a form wizard.

Authors are given access to create or edit documents in the database.

A special supplement that allows access to materials of the encyclopedia is given to registered authors. Anonymous and free access to documents of certain types from global networks is also possible.

Access to the contents of electronic encyclopedias is possible by a special browser providing viewing, searching and reception of necessary materials.

Documents forms.

Document forms represent a standard of hierarchical representation of objects and concepts descriptions and provide:

- multiple-valued and multilevel classification of documents, making their internal contents understood;
- creation of a hierarchy of descriptions varying in completeness of concepts;
- opportunity to accommodate and to use interactive documents with professional bases of knowledge as fields content.

The set of used forms creates a special base. Any form is a sequence, which consists of:

- separate forms, obligatory in the documents structure;

- collections of variants of the forms, allowing to include in the structure of the documents only one variant to have the opportunity of different specification of separate forms sections;
- serial forms which can be filled with any amount of copies in any concrete document;
- forms of the indexes, allowing to organize connections with other documents of an electronic collection or parts of such documents.

Separate values from created documents are placed in the elementary forms, named as elementary fields. Such values belong to standard or designed types, which are determined by the help of sets of logical and structural restrictions.

VEDA makes it possible to make a description of the value structure for the elementary field of the documents, containing separate phrases or groups of phrases of standard structure, by means of the sequential-parallel (s-p) networks with marked vertices and edges and references on vocabularies and grammar rules. This makes it possible for representation and processing of semantics for the sentences in limited natural language constructions.

The special class of forms allows us to create documents which fields of values contain fragments of professional knowledge bases. These are created in one of the described and stored models of representation, focused on the application of knowledge processing programs with the aim to solve the **user's** problems. Such documents are interactive, assuming the organization of users access is multilevel dialogue, providing an opportunity of a formulation and specification of the user problems at the beginning of the document processing, and an explanation of the received results on the final stage of the dialogue.

The opportunity to add knowledge representation models and methods of their processing to the system of standard and designed types of fields' values is stipulated in VEDA. The special tools that allow creating and applying knowledge bases for model of predicate rules with exceptions and specifications are developed for the contemporary VEDA version.

Interactive documents with predicate rules bases as elementary fields content allow carrying out consultation, diagnostics, forecasting and testing with the help of special procedures of knowledge processing [2].

Documents classification.

Two approaches on structure and on a subject domain of documents are used for document classification in the VEDA system:

Classification of documents by structure is based on an affinity of structure for the used forms. It generates the relation *general - special*, forming a lattice on the set of the forms determined for each concrete area of the application and is defined by rules:

1. If T1 and T2 – are elementary forms, then T1 is called more general iff for sets DT1 and DT2 of values of unique fields of such forms inclusion DT1 in DT2 take place;
2. If forms $T1 = (V_1, \dots, V_n)$ and $T2 = (R_1, \dots, R_m)$ are given, then form T1 is more general, than T2 iff for different elements of list R_1, \dots, R_m there exist different more general elements from V_1, \dots, V_n .

The application of such a classification allows to realize different functions of processing the documents of close and similar structure: to search the necessary information, to maintain the universal access to the documents series of developing structure, to carry the documents content from one form to another.

Classification of documents by a subject domain is connected to the allocation of the documents to the system of areas of its assignment or application. For each such area the collection of assigned forms for documents used in it is set. It is possible to use common forms for different areas of their application. The documents created on the basis of such forms receive multiple-valued classification.

Document visualisation.

The visualisation of the documents is connected to the necessity of granting users the information, taken from documents in the required form and sequence. For such a purpose any form is assigned with one or several scripts of visualisation. For example, for the document, focused on user reading, visualisation is connected page by page on a multilevel hierarchy to display its contents.

For interactive documents visualisation stages are connected to the realisation of a dialogue with the users, consisting of specification of conditions and parameters of decided problems; requirements for the received decision, and also distribution and explanation of transmitted decisions and recommendations. Thus the processing of a document and its contents appear closed for the user.

Semantic net for documents space.

A network of semantic connections between articles represents a knowledge area of the encyclopedia. Every connection is created by a compound reference: in the documents and in parts of other documents within their own structures. Tools for a semantic net of encyclopaedia processing will be developed for future version of VEDA.

Search by association.

The search of necessary information is carried out on the basis of inquiries represented as search forms, created for separate forms or form groups of documents. They contain the description of the possible contents of the found documents: connections between documents, patterns or conditions for values of separate fields, sizes of a degree of concurrence between documents and forms.

Identification of the materials of the encyclopedia and the inquiries of the users is carried out by construction of associations for search forms and documents contents. Associative searches provides allocation of relatives on structure or the contents of documents with the help of an establishment of similarity between documents and search samples on the basis of structural, symbolical and functional unification mechanisms and on estimation of a degree of concurrence [3].

The contemporary version of the VEDA system forms and processes search samples on the basis of the following rules:

- the inquiry has a name and weight, as value from a range $[0,1]$ which determines its importance;
- the inquiry happens to be the elementary, represented by conditional expressions, or compound or complex, represented by lists of subinquiries included in it, forming a hierarchy with elementary inquiries as final elements;
- every inquiry can be declared obligatory for the general inquiry it is contained in;
- compound inquiries contain the common condition as restriction on the weighed sum for carried out subinquiries;
- elementary inquiries are made as the logic expressions made of comparisons for values of fields of documents and logic operations.

Feasibility of elementary inquiry is equivalent to the validity of the appropriate logic expression. The complex inquiry is feasible if the weighed sum for executed subinquiries satisfies the common condition and all obligatory subinquiries are executed.

The expanded version of the VEDA system will allow to set structural descriptions of values for separate fields and groups of fields of documents as fragments of semantic networks and to apply them at the construction of inquiries. Thus there is an opportunity of inclusion in search samples of the network patterns supposing the use of the unification process, both for values of fields of documents, and for semantic connections inside and between fields in them.

Regular user applications

The encyclopedias in the VEDA environment allow organizing services of regular user applications.

Such applications consist of the inquiries containing data of user registrations, parameters of search space, search samples and conditions of their application, conclusions of information search, time and spatial parameters of service.

The service of regular information tasks is carried out in the form of dispatches of the found documents, fragments of documents or references on documents.

The set of regular inquiries is placed in a special base, which is processed by a stream of documents, placed in a storehouse. Any such document cooperates with the base of inquiries, allocating such inquiries for what it is carried out. Internal formats of the base of inquiries for the VEDA system allows to organize the parallel processing of inquiries by taking into account logic dependencies between them and their fragments.

Tools of access to collections of documents from global networks.

The standard way to access documents of the electronic encyclopedia is to use a special Web-browser. It provides an opportunity of viewing contents, tables of contents and makes the documents placed in the storehouse visible in the HTML format. Access for authors and owners of the documents in the encyclopaedia is provided by a special supplement of the forms and by the document wizard. Creation of regular information inquiries is also carried out by the special supplement.

Approbation of VEDA system.

Today the VEDA system is used to make an encyclopaedic collection of professional knowledge in the field of theoretical foundation of selection. Experts of the state research institute have designed five base structures of descriptions of documents: term's explanations, concepts, techniques, sources and field tests.

Design of the structures of the documents was carried out in two consecutive stages:

- From the top downward, when four allocated classes of documents systems of subforms were consistently specified;
- From below upwards, when hierarchical structures of forms, elementary forms, fields of values and forms indexes, plus their values were projected.

New projects for the VEDA system.

Scientists of the Kuban University now prepare materials for electronic encyclopedias in several fields of knowledge. One of them is about Cossacks of Northern Caucasus. This project will be using the VEDA system.

Other electronic encyclopaedias, being planned, are digital projects on wildlife management and ecology in the Krasnodar territory.

References.

1. K.I. Kostenko, B.E. Levitsky The Concept of documentary resources for system of library and information establishments of Krasnodar territory. The Second Intern.Conf. ISP, St.-Petersburg, 2000, p. 371.
2. K.I. Kostenko, S.V. Nazarenko Intellectual technologies for libraries in the Internet. The Second Intern.Conf. ISP, St.-Petersburg, 2000, p. 372-373.
3. M.M. Denisov, K.I. Kostenko Representation and service of regular queries in storehouses of documents in system VEDA The Intern.Conf. TELEMATICA2000, St.-Petersburg, 2000, p.28-29.