

# Thesaurus of the Discipline on Information Security Management System

Niyazova R.S., Sexenbayeva A.K.,  
Baigabyl U.D., Iskakova D.B.  
L.N.Gumilyov Eurasian National University  
Astana, The Republic of Kazakhstan  
Email: rozamgul@List.ru, aiisha\_08@mail.ru,  
baigabyl\_uldana@mail.ru, iskakova.dariya@bk.ru

**Abstract**—This paper, the technique of construction of an WordNet type lexical-semantic database in Kazakh language is offered, questions of this database integration with the same type of databases and its representation in the OWL format are described.

**Keywords**—lexical-semantic database, thesaurus, OWL.

## I. INTRODUCTION

The primary concern of this research is to build Kazakh-English WordNet of the discipline on information security management systems, which includes lexicons of Kazakh and English languages. The development of such a variant of the dictionary consists of two stages - constructing of Kazakh WordNet with describing of the vocabulary of the Kazakh language and combining of the resulting WordNet with the latest implementation of The Princeton WordNet by the use of ILI.

For reviewing and editing of dictionary entries of WordNet and hierarchy of their relationships trees of hyponymies (species of the genus relations) and meronymies (part-whole relationships) are built. The editor is used to "finish" the processing of the dictionary.

## II. STAGES OF CONSTRUCTING

The practice of designing and constructing of the dictionary has showed that editing of a specially prepared text files and a set of utilities for making changes to the database are the most effective. For each stage of building a dictionary formed a set of files and their means of treatment. These sets of files and their means of treatment are formed for each stage of building of the dictionary.

The basic structural unit forming a dictionary paper of WordNet is a synonymous series - "synset"- combining tokens with a similar meaning.

Each synset represents a certain value, a concept of a language. To clarify a value of a synset its interpretation and usage examples of tokens of a synset in some context are determined. Each synset represents a certain value, a concept of a language. To clarify a value of a synset its interpretation and usage examples of tokens of a synset in some context are determined. Information retrieval thesaurus (IST) - a controlled vocabulary of terms of a natural language,

indicating the relationships between the terms and intended for information re-trieval.

Mains relation of thesauruses: Synonym, antonym, hyperonym, hyponym, meronym.

EuroWordNet. Resource the WordNet, only developed for the English, called the world's great interest in the development of such re-sources for dozens of other languages.

Developing WordNet-s for the different languages within EuroWordNet project includes two phases. In the first phase (1996-1999) EuroWordNet began as an EU project, with the goal of developing wordnets for Dutch, Spanish and Italian, and to link these wordnets to the English wordnet in a multi-lingual database. In 1997, the project was extended to include German, French, Czech and Estonian. EuroWordNet (<http://www.hum.uva.nl/ewn>) was completed at the end of 1999.

There was a need to do a serious choice in the project: whether to strive for the development of a language-independent structure, with which it is necessary to compare the units of each language, or maybe it needs to have a common system of synsets - a new unit in the hierarchical network can be enabled, if at least one language from considered has a token or a steady turnover with a value.

According to the adopted in the draft decision, each WordNet must retain the specificity of its language. Thus each WordNet should contain references to the values of the English WordNet that allows to compare WordNets, detect inconsistencies in their constructing and see the differences in the designs of language systems. At the same time a small top-level ontology was created in the framework of the project, to which must be attributed to each created WordNet.

Authors of EuroWordNet emphasize the distinction between a resource of wordnet classes as linguistic ontologies and formal ontologies. Linguistic ontology must reflect the relationship between lexicalized words and expressions of the language, for example, to describe what words can be used to replace the text in the word spoon (spoon) - object, tableware, silverware, merchandise, cutlery.

Thus, WordNet- is a network of language-specific lexicalized units (as opposed to formal ontologies, which represent the data structure with a formal definition).

Basic alleged use of WordNet - the prediction of a possible

replacement of lexical units in the text for the purpose of information retrieval, text generation, machine translation, word-sense disambiguation.

Given the difficulties that arose in the application of WordNet applications, European developers have proposed a number of significant innovations in the structure created by WordNet. A large class of these changes is concerned with describing of the relations between synsets, which can be divided into the following groups:

- Ascription of additional attributes to existing relationships;
- Introduction of relations between the different parts of a speech;
- Introduction of additional relationships between words (synsets) of the same part of a speech.

RDF - language representing information about the WWW resources. In particular, RDF is used to represent the metadata associated with web resources, such as "title", "author", "date last modified." But RDF can be used also to provide information about the resources of the "second type", to which it is possible to refer only (or to identify on the Network through URI), but it is impossible to get to them access through the Network directly.

It can turn out that in certain cases for control of meta data it is enough to use XML and XML Schema (or generally to be restricted to a sub-element of HEAD of the HTML element). But this approach is poorly scaled: in case of increase in volume of meta data, complication of their structure control of the meta data constructed on the basis of XML Schema becomes the labor-consuming task for which decision RDF is intended.

Among the semantic relations in the dictionary of the Kazakh thesaurus WordNet the following nouns are set: Hyperonym, Hypo-nym, Meronym, Holonym, Ontonym, Homonym, Synonym. Having created the new project, we realize it further in the environment of Protege.

The last description of the standard language – the language OWL, World Wide Web Consortium (W3C).

Protégé OWL describes not only concepts or concrete objects. Protégé OWL has a great number of operators, for example intersections, merges, denial and many other things. Mostly it is made on the basis of an informal logical model.

The main goal of our work consists in development of the Kazakh language WordNet. Such option of development consists of two stages - the subject area of the Kazakh language WordNet - lexicographic system and the qualifier of system. To program the project built in the programming environment Protege as a basis need to take a thesaurus ontological model, build in Python programming environment. The initial window of building of lexical analysis is WordNet Kaz. When searching for words we can see the full lexical analysis.

### III. CONCLUSION

The considered system intends for creation and editing of a wide class of thesauruses and the close to them structures.

Implementation of a set of interfaces to these systems allows to use them as independent applications – the lexicographic WordNet system and system of qualifiers, and to include them in composition of more difficult systems.

### REFERENCES

- [1] C.Fellbaum (ed.), WordNet: An Electronic Lexical Database, MIT Press, 1998.
- [2] P.Vossen. Building a multilingual database with wordnets for several European languages. <http://www.ilc.uva.nl/EuroWordNet/>
- [3] A.Horak, P.Smrz, P.VisDic – Wordnet Browsing and Editing Tool. P.Sojka, K.Pala, P.Smrz, C.Fellbaum, P.Vossen (Eds.): GWC 2004, Proceedings, pp. 136–141. Masaryk University, Brno, 2003
- [4] OWL Web Ontology Language. Guide. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/2004/REC-owl-guide-20040210>
- [5] OWL Web Ontology Language. Reference. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/2004/REC-owl-ref-20040210>
- [6] Ukhonogov A. M., Yablonskiya S. A. University of ways of messages of Peter-burg, zao Russikon x system.

### ТЕЗАУРУС ПО ПРЕДМЕТУ СИСТЕМЫ УПРАВЛЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Ниязова Р.С., Сексенбаева А.К.,  
Байғабұл Ұ.Д., Искакова Д.Б.

В работе предлагается методика построения лексико-семантической базы данных типа WordNet по предмету системы управления информационной безопасности на казахском языке. Описываются вопросы её интеграции с другими базами данных WordNet и представление в формате OWL.

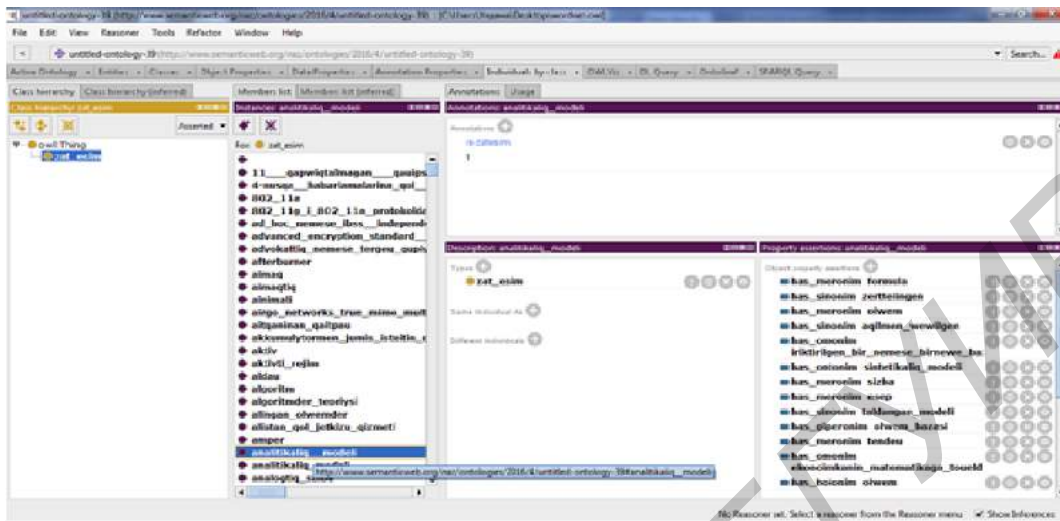


Figure 1. Writing the words of an individual

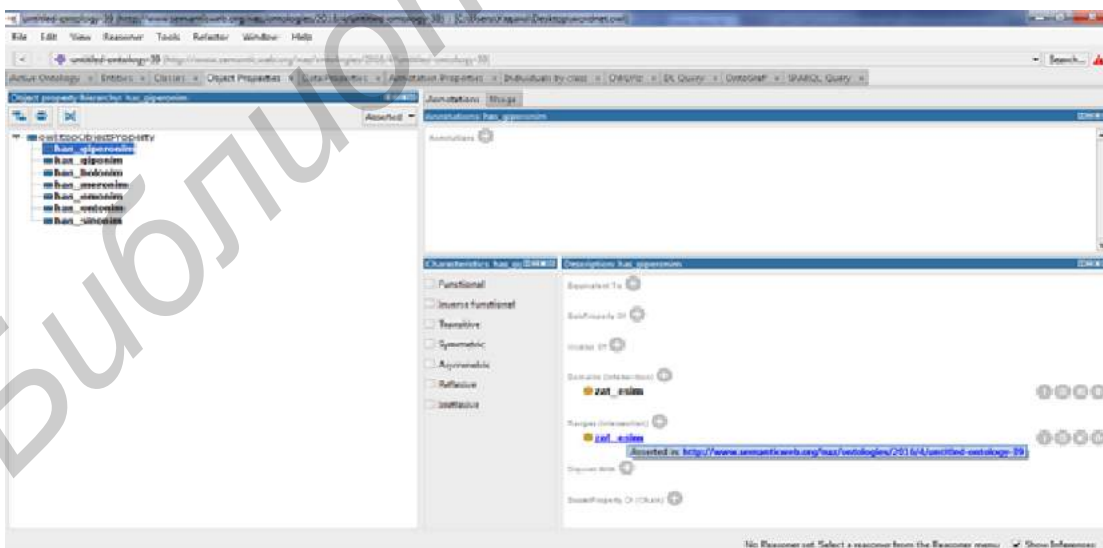


Figure 2. Window of properties of an object. (Object PropertiesTab)



Figure 3. Showing of object properties

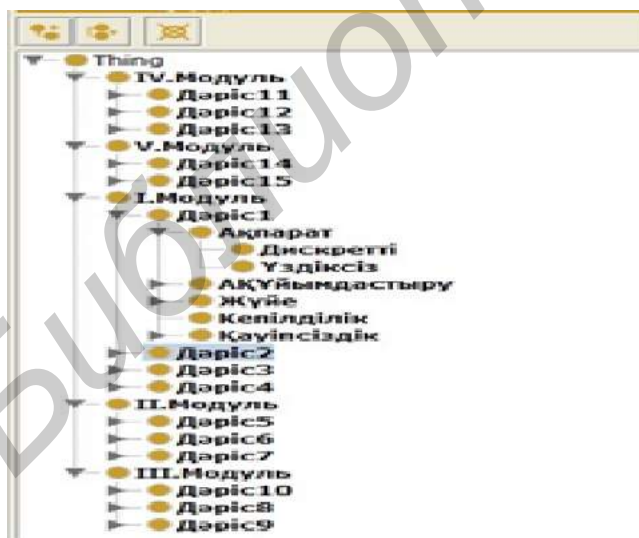


Figure 4. Division into classes and internal classes

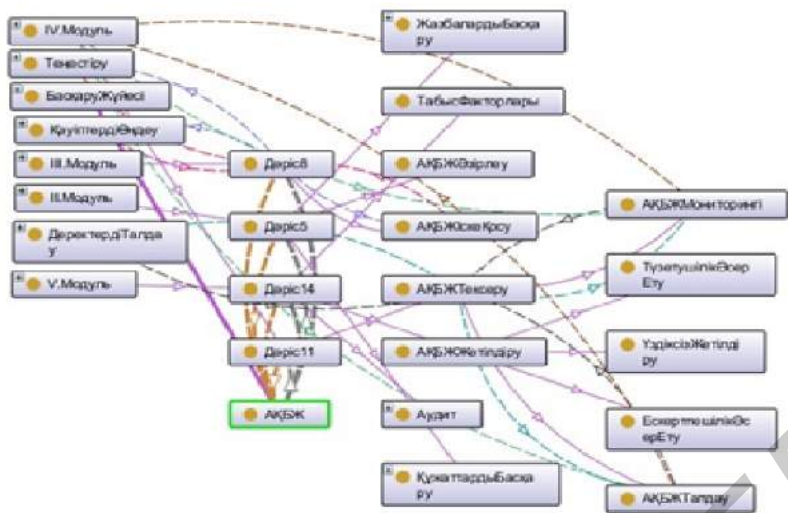


Figure 5. Relationship between classes and modules of concepts of ISMS

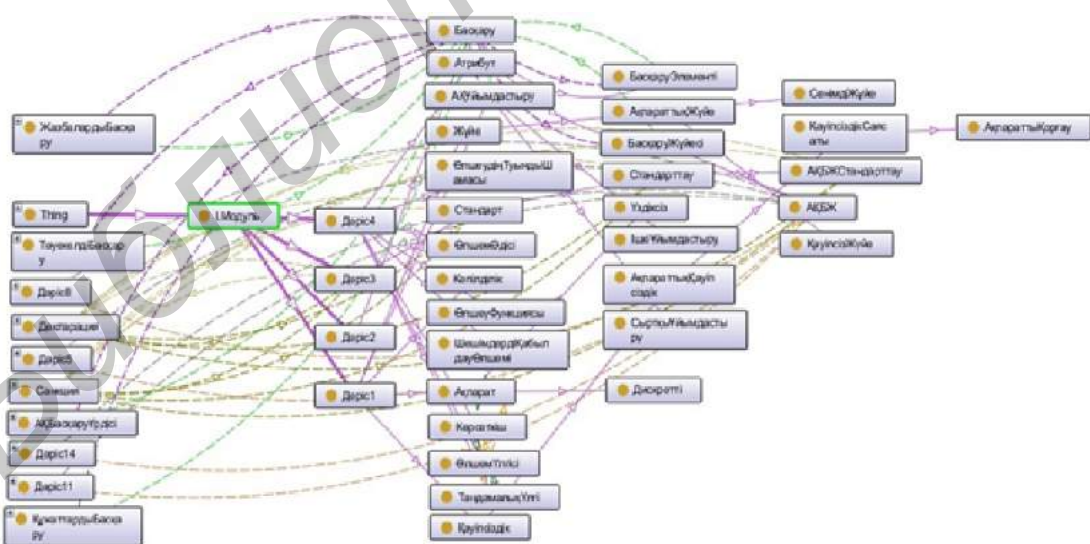


Figure 6. View of general ontological model 1

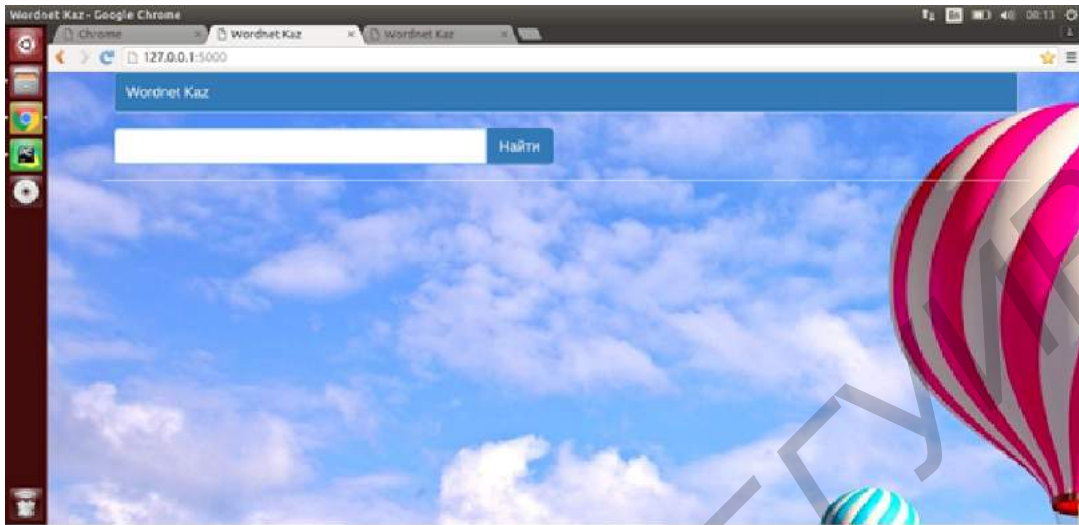


Figure 7. Showing of the Results of Wordnet Kaz

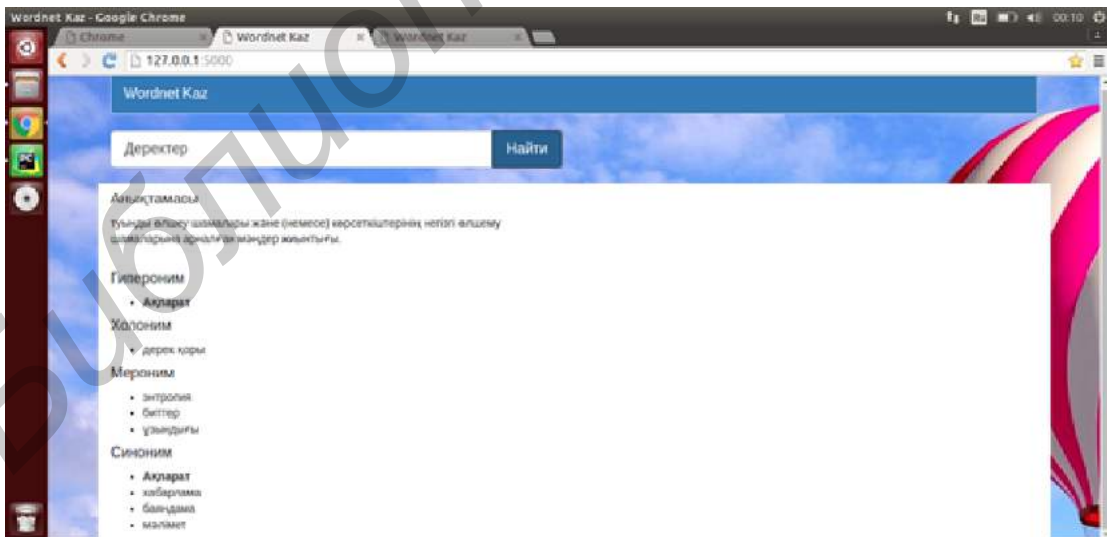


Figure 8. The main form