Thesaurus and Domain Ontology of Ecoinformatics

Boryana Deliyska[1], George Manoilov[2], Peter P. Manoilov[3]

[1] *Department of Computer Science and Informatics, University of Forestry, Sofia, Bulgaria*
[2] *Product and Service Manager in Imbility Ltd., Sofia*
[3] *Denima Inc, Sofia*

Abstract. Ecoinformatics is interdisciplinary science and as a part of informatics, environmental science and ecology uses methods and terminology of informatics and many natural sciences. The current work comprises thesaurus and domain ontology of ecoinformatics design and implementation. In the article ontology of ecoinformatics is developed in the form of thesaurus and OWL code. It is domain ontology with layered structure consisting of syntactic, internal and external link layers. For the purpose as a text corpus existing multilingual dictionary of ecoinformatics is used. Concepts about its structure and relationships with other scienses are included. The thesaurus and ontology building is preceded by ecoinformatics taxonomy development. The thesaurus and domain ontology are base for building application ontology of ecoinformatics course syllabus and for information systems design and implementation.

Keywords: ecoinformatics, domain ontology, thesaurus, taxonomy

1. INTRODUCTION

In contrast with the essentially spatial character of geo-ontologies, eco-ontologies are fundamentally temporal in character (Fonseca F., Martin J., Rodriguez A. 2002).

In [Deliiska B. 2007b] the nature and problems in ecoinformatics are researched. But the links between basic concepts in the domain and to other sciences are not clarified completely. Formal description of these links would be useful for knowledge management in the domain.

The purpose of this work is development of thesaurus and ontology of ecoinformatics helping future educational courses and information system design and exploitation.

The paper is structured in five sections. The next section examines external links of ecoinformatics with other scientific domains. In the third section the process of creation of ecoinformatics thesaurus and ontology is described. The fourth section takes an account of thesaurus and ontology application. At the end conclusions and intention for future work are represented.

2. EXTERNAL LINKS IN ECOINFORMATICS

Here two types of relationships between ecoinformatics and other scientific domains are examined. The first type assumes the availability of well-developed ontologies of nearest domains – informatics, bioinformatics and geoinformatics, and is represented by simple schema (fig.1). For example, Gene Ontology [Ashburner M. et al. 2000] is among the most popular ontologies of bioinformatics and computational biology. Ontologies about geoinformatics, GIS and geography in *[Gutierrez L. 2006; Fonseca F., Egenhofer M., Agouris P., Câmara G. 2002; Smith B. 2001]* are described.

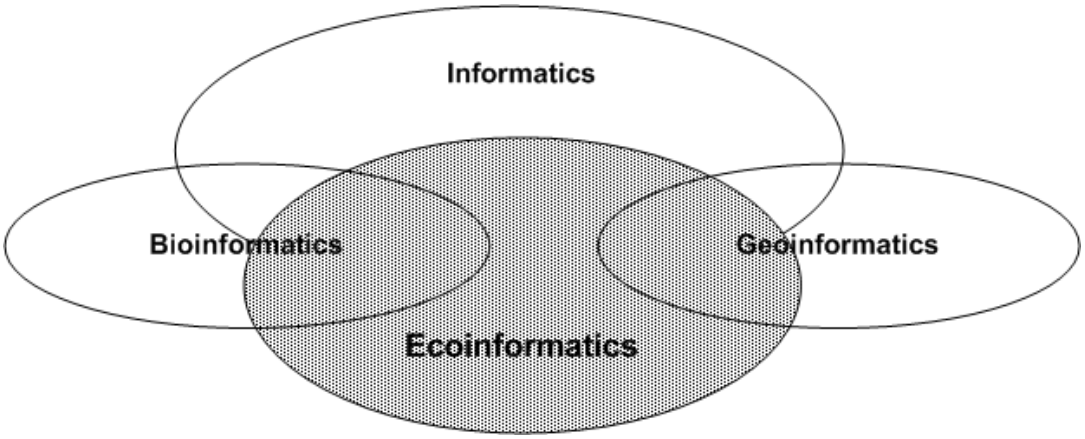

Fig. 1. Common schema of ecoinformatics ontology external links

Other domain ontologies in well-known ontology libraries as OntoSelect [http://olp.dfki.de/OntoSelect/index.php?mode=home], SchemaWeb [http://www.schemaweb.info/schema/BrowseSchema.aspx], DAML [http://www.daml.org/ontologies], Alexandria Digital Library [http://www.alexandria.ucsb.edu/], SWOOGLE [http://swoogle.umbc.edu/] etc. are released.

Because more of these ontologies are inadequate for the concrete purpose, in the second case detailed schema of relationships is developed (fig.2) like that in the domain of geoinformatics [Delijska B. 2007a].

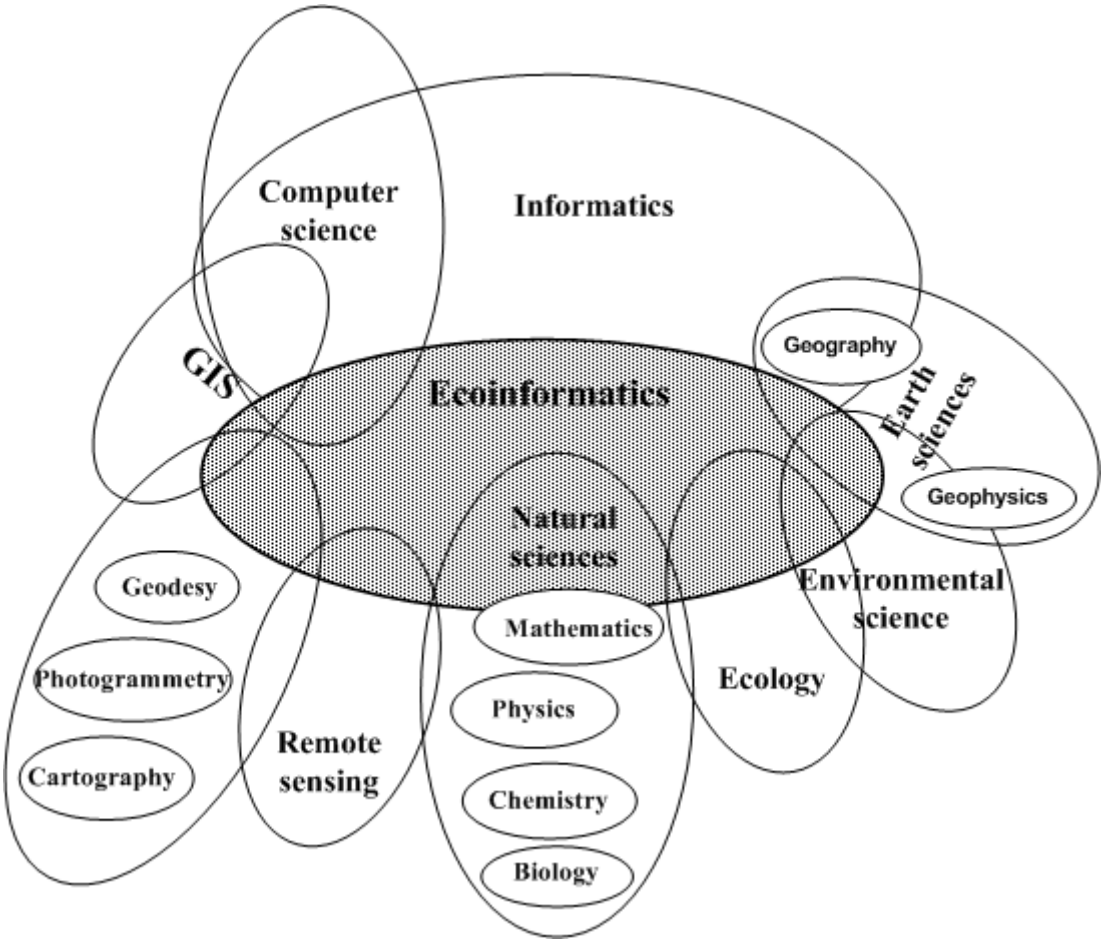

Fig. 2. Detailed schema of external links in ecoinformatics ontology

Obviously, in ecoinformatics are used methods, algorithms and terminology of informatics, computer science, natural sciences, earth sciences, geodesy, remote sensing, ecology, environmental sciences etc.

3. THESAURUS AND ONTOLOGY OF ECOINFORMATICS

Ontologies in the area of ecoinformatics have hard semantic heterogeneity, due to different conceptualization of the connected scientific domains by different researchers. There are no commonly agreed methodologies about the development of shared and consensual ontologies [Gómez-Pérez, Benjamins 1999] but all of them comprise the following consecutive stages:

- domain conceptualization and controlled vocabulary building;
- establishment of hierarchical relations between concepts and taxonomy design;
- definition of non-hierarchical relations, facets and other ontology elements and its formalization;
- verification;
- ontology publication with a view to its reusability.

Development of ecoinformatics taxonomy is based on existing web-based glossary of ecology [Deliyska, Manoilov 2003], extended with the terminology extracted from investigations in [Deliiska 2007b] and [Michener 2006].

3.1. Ecoinformatics Taxonomy

An excerpt of the upper levels ecoinformatics taxonomy in fig. 3 is shown.

Top class *Ecoinformatics* comprises subclasses *Theoretical Ecoinformatics*, *Applied Ecoinformatics* and *Ecoinformatics Application*. In the next level class *Theoretical Ecoinformatics* has subordinate classes *Mathematical foundations of Ecoinformatics*, *Ecoinformatics Methodology* and *Environmental Metadata Standardization*.

Class *Ecoinformatics Methodology* contains subclasses of methods, used in ecoinformatics problem solving: ecosystem modeling, analysis, synthesis, prognostification etc. [Deliiska 2007 b].

Class *Mathematical foundation* includes concepts of discrete mathematics, differential calculus, artificial neural networks, cell automata, graph theory, set theory, formal logics, artificial intellect, genetic and evolutional algorithms and programming, mathematical topology, system theory etc., defined in an ontology of mathematics.

Subclasses of *Applied Ecoinformatics* are *Environmental Information System* (*EIS*) and *Environmental Expert System*. On the next level subclasses of *EIS* are *People*, *Environmental data*, *EIS hardware*, *EIS software*, *EIS organization*, *EIS product* and *EIS functionality*. Class EES has similar subclasses excluding class *Eco-knowledge base*.

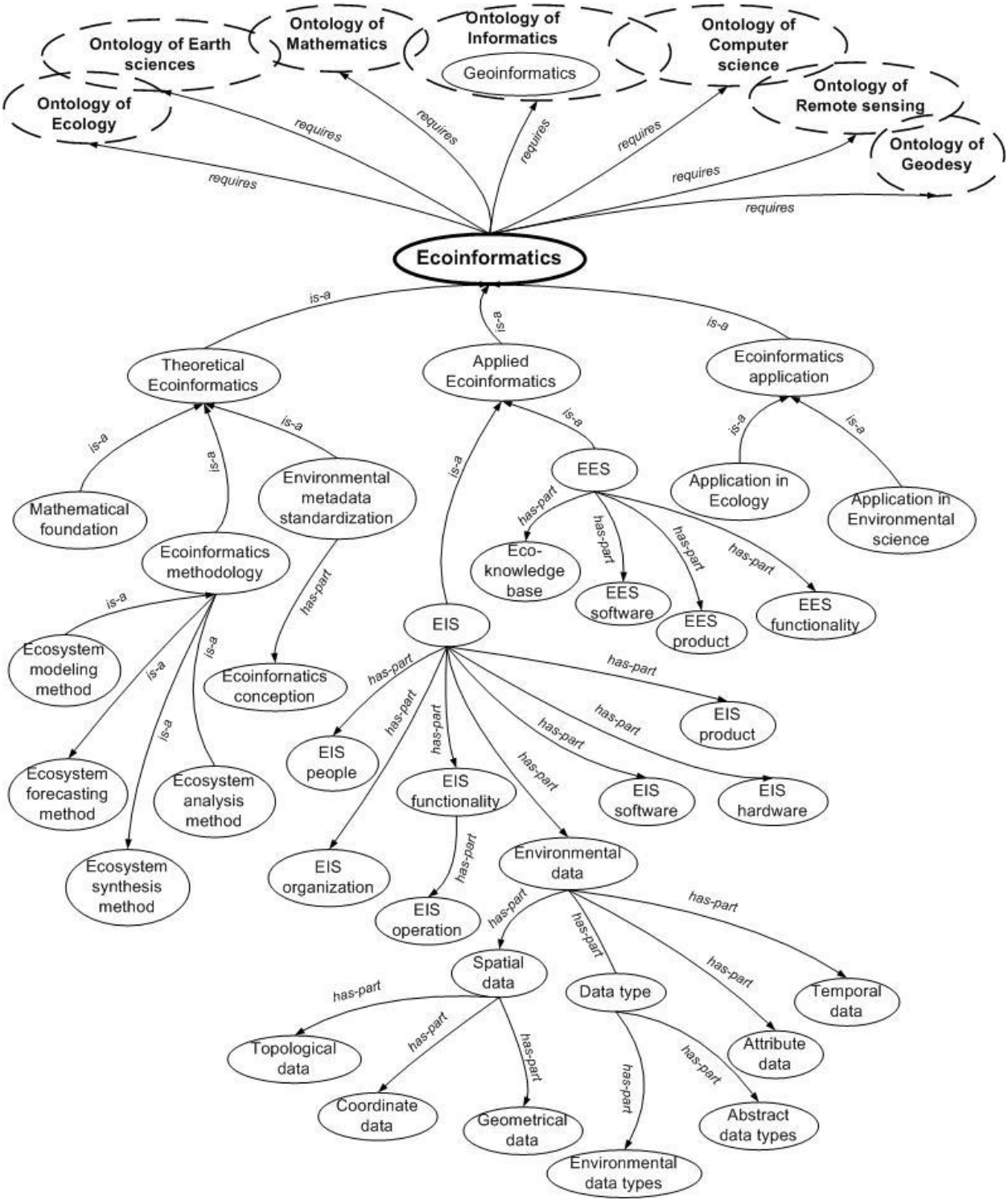

Fig. 3. An excerpt of upper levels of ecoinformatics taxonomy

3.2. Thesaurus

The thesaurus as a form of ontology formalization, represents full taxonomy, supplemented with non-hierarchical relations, non-preferred terms (synonyms and lexical variants), definitions etc. Between ecoinformatics concepts non-hierarchical location and topology relations exist (for example, *is-located-in, is-boundary-for,*

intersects etc.). Five-lingual (English-Bulgarian-German-French-Russian) web thesaurus with leading English is created. On fig.4 a screen of its database is shown.

Note: The number and type of languages do not affect the processing algorithm and could be easy changed.

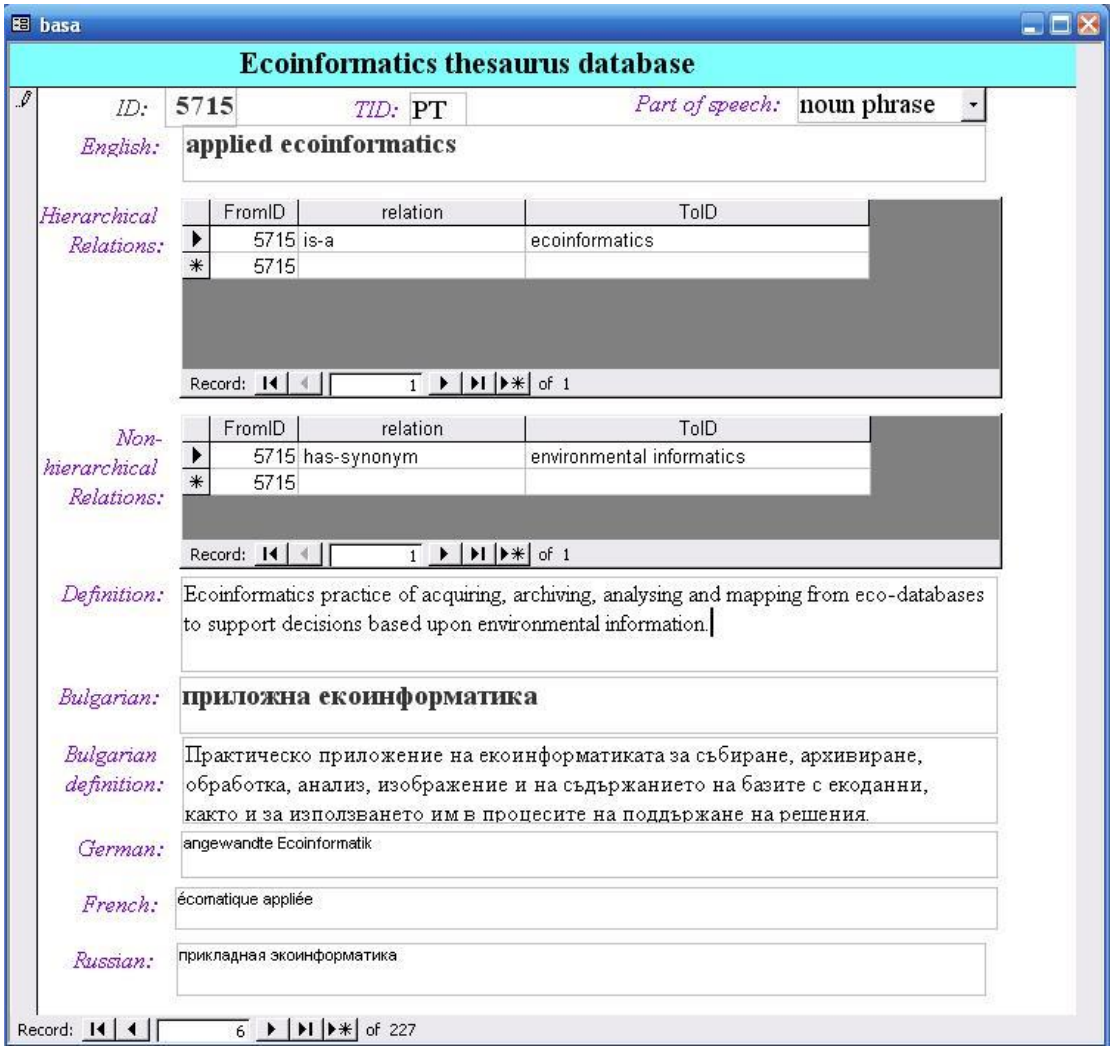

Fig. 4. Screenshot of ecoinformatics thesaurus database

The thesaurus includes about 300 lexical units and is available on http://www.e-forestinformatics.org/projects/Ontologies/Ontology_ecoinformatics/ontology_bg.html. The interface page offers two kind of queries – by full or part term name or by group of terms, beginnig with selected letter.

3.3. Ontology

An important reason for converting the thesaurus into a formal ontology language is the necessity of its machine-readability and reuse. In (van Assem et al. 2004) a method for conversion of thesauri into ontology code is proposed, containing four steps: (0) a preparatory step, (1) a syntactic conversion step, (2) a semantic conversion step, and (3) a standardization step.

In this work a method developed in [Deliiska 2007a] is used. In accordance with this method, the ecoinformatics ontology is formalized in *OWL* by converting the thesaurus

with VBA application. The migration from a thesaurus database to OWL ontology knowledge base is semi-automatic because after conversion some manual adjustments and refinements in the Protégé editor are necessary. Then, syntactic verification by W3C Validator [http://www. w3.org/RDF/Validator]. Further the ontology code in Protégé editor is open where external links are added to:

б) част от понятията, използвани в онтологията по екоинформатика, са включени в някои онтологии, публикувани в библиотеката на семантичната търсачка SWOOGLE *[http://swoogle.umbc.edu/]. Тези онтологии са за време и пространство [http://ecoinformatics. uvm.edu/dmaps/growl/TimeAndSpace.owl#], статистически анализ [http://cvs. ecoinformatics. org/cvs/cvsweb.cgi/seek/projects/kr-sms/OWLOntologies/StatisticalAnalysis.owl#], основни мерни единици [http://cvs.ecoinformatics.org/cvs/cvsweb.cgi/seek/projects/kr-sms/OWL Ontologies/MeasurementBase.owl#], модели в екологията [http://cvs.ecoinformatics.org/cvs/ cvsweb.cgi/seek/projects/kr-sms/OWLOntologies/EcologicalModels.owl#], мрежи [http://cvs. ecoinformatics.org/cvs/cvsweb.cgi/seek/projects/kr-sms/OWL Ontologies/Networks.owl#] и за конепции в екологията [http://cvs.ecoinformatics.org/cvs/ cvsweb.cgi/seek/projects/kr-sms/ OWLOntologies/EcologicalConcepts.owl#].*

Към тях се добавят пространствата на имената на онтологиите за областите, свързани с екоинформатиката, разработени тук и изложени в т.3.2 на настоящия раздел.

http://www.e-forestinformatics.org/projects/Ontologies/ontologies_en.html.

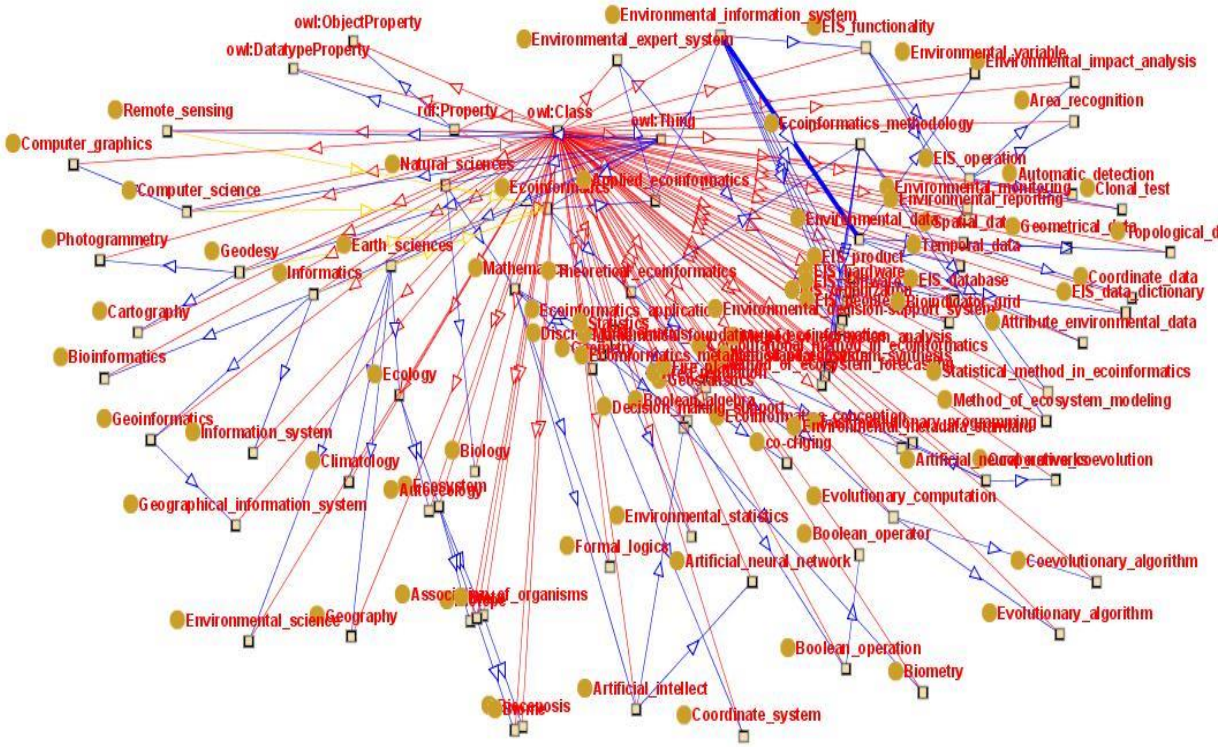

Fig. 5. Graphical representation of ecoinformatics ontology

4. THESAURUS AND ONTOLOGY IMPLEMENTATION

5. CONCLUSIONS

The proposed ontology of ecoinformatics is based on our own experience in research and teaching in the area. Moreover our long practical skills in dictionary compilation and information systems development were essential during the ontology building.

REFERENCES

[1] Ashburner M. et al. 2000 Gene ontology: tool for the unifcation of biology. The Gene Ontology Consortium. Nature Genetics, pp.225-29

Deliiska B. 2007a Thesaurus and domain ontology of geoinformatics, J. Transaction in GIS, ISSN**:** 1361-1682, vol 12 issue 4 (11), 2007, pp.637-651

Deliiska B. 2007b Survey of Ecoinformatics, Balkan Ecology, vol.10, 2007

Gómez-Pérez, A.; Benjamins, R. 1999. Applications of Ontologies and Problem Solving Methods. AI Magazine. Vol. 20[1] B. Spring, 199-122

Fonseca F., Martin J., Rodriguez A. 2002 From Geo to Eco-Ontologies, Proceedings of the Second International Conference on Geographic Information Science, p. 93-107, ISBN:3-540-44253-7

Michener, W.K. 2006 Meta-information concepts for ecological data management, Ecological Informatics 1, pp. 3-7

van Assem M., Menken M.R., Schreiber G., JWielemaker J., Wielinga B. 2004 A method for converting thesauri to RDF/OWL, in the Third International Semantic Web Conference (ISWC), Hiroshima, Japan

Williams R., Martinez N., Golbeck J. 2006 Ontologies for ecoinformatics, Journal of web semantics, Web Semantics: Science, Services and Agents on the World Wide Web 4 (2006) 237–242