Marinos Kavouras & Margarita Kokla
Department of Rural and Surveying Engineering
National Technical University of Athens
9, H. Polytechniou Str., 157 80 Zografos Campus, Athens - Greece
Tel: 30+1+772-2731/2637, Fax: 30+1+772-2734
mkav@survey.ntua.gr, mkokla@survey.ntua.gr

ONTOLOGY-BASED FUSION OF GEOGRAPHIC DATABASES

ABSTRACT

Effective communication and smooth interaction between different sources of geodata require a method for sharing and integrating different ontologies. The present paper proposes a methodology for information organization and semantic integration, in order to provide reuse of data between heterogeneous geographic information systems.

The methodology is founded on Formal Concept Analysis, a theory of concept formation and conceptual classification. The integration of multiple geospatial categorizations, which exhibit differences in spatial and thematic resolution, allows the creation of an ontology for the geospatial domain. Furthermore, the methodology and the integration process can be utilized to build a multi-scale, multi-context database from different geographic categorizations, which represents information at different levels of detail and different application contexts. The final integrated geographic ontology is demonstrated and queried using an Ontology Browser.

Key Words: geospatial ontologies, formal concept analysis, semantic integration, multi-scale, multi-context.

INTRODUCTION

Interoperability aims at the development of mechanisms to resolve any incompatibility and heterogeneity and to ensure access to data from multiple sources. The dynamic interaction of different applications requires not only the technical support for the exchange of data, but the preservation of the underlying semantics as well. However, although, the technical aspect of data exchange is developed successfully due to advances in information technology, issues related to the semantic aspect need further examination.

Sharing geospatial data is difficult due to diverse conceptual schemata and semantics. Indeed, different interpretations of geospatial data encoded in different databases cause heterogeneities between them. Heterogeneities between different databases can be classified according to three major categories (Bishr, 1998):

- **Syntactic Heterogeneity** is caused by different logical data models (e.g., relational vs. object-oriented) or due to different geometric representations (raster vs. vector).
- **Schematic Heterogeneity** occurs because of different conceptual data models (e.g., objects in one database considered as properties in another, different generalization hierarchies).
- **Semantic Heterogeneity** raises most information integration problems. It occurs because of differences in meaning, interpretation or usage of the same or related data. Semantic heterogeneity is divided to:
- naming heterogeneity (homonyms and synonyms), and
- cognitive heterogeneity: different conceptualizations e.g., class definitions or geometric descriptions

The main causes of semantic heterogeneity are the differences in the conceptualization of geographic data in conjunction with their complexity. Different geographic categorizations (differences in application context and levels of detail) pose a semantic problem when geospatial applications have to be integrated.

In order to achieve semantic interoperability between different geospatial applications, a commonly accepted theory for the formal definition and representation of the semantics of geospatial knowledge would be ideal. This theory would provide the basis for the formal representation of geographic entities with regard to their structure and semantics. However, besides this ideal, long-term goal, there is also immediate priority to develop suitable methods and tools to formalize geospatial concepts and relationships encoded in existing databases, in order to enable database fusion.

Ontology (Guarino, 1998; Smith, 1998; Sowa, 2000), as studied by both philosophy and AI community, is considered an important contribution towards the achievement of interoperability. For philosophy, Ontology is defined as the study of the categories of things that exist or may exist in some domain. For the AI community, the notion of Formal Ontology is used to denote a collection of concept and relation types specified by axioms or definitions stated in a formal language and organized by the type-subtype relation. For the geographic domain, ontologies play an important role in defining the semantics of geographic information and facilitating information integration between different databases.

The methodology presented in this paper focuses on the formalization of geospatial concepts and relationships using Formal Concept Analysis (Wille, 1992, Ganter and Wille, 1999) and the integration of multiple geographic categorizations, which exhibit differences in application context and thematic resolution. These objectives facilitate geographic information sharing between different organizations and for different purposes.

The remainder of the paper is organized as follows. Section 2 introduces the proposed methodology. More specifically, the integration of different geographic categorizations is described in Section 2.1, whereas the utilization of the integrated geographic categorization for building a multi-scale, multi-context database is presented in Section 2.2. Section 2.3 describes the development of an Ontology Browser, which is used as a tool to create and manipulate ontologies. Finally, an overall evaluation of the proposed methodology is presented in Section 3.

PROPOSED METHODOLOGY

Integration of Different Categorizations

In order to demonstrate in a comprehensive fashion the application of the proposed methodology, a running example is used involving the integration of three independent classification schemata:

- The hierarchical CORINE Land Cover nomenclature (CORINE Land Cover-Technical Guide, 1994) for scales 1:100,000–1:1,000,000.
- The DIGEST nomenclature for geographic objects (DIGEST Standards Specification, DGIWG, 1997), addressing a variety of scales.
- The classification used by the Hellenic Mapping and Cadastral Organization (Technical Specifications of the Greek Cadastre, HEMCO, 1996) to record land use characteristics referring to scales 1:1,000–1:5,000.

The process of integrating multiple categorizations is divided in two main steps: *Semantic Factoring* and *Concept Lattices*. Semantic Factoring is the process of analyzing-decomposing the categories of the original categorizations into a set of fundamental categories. At this step, it is necessary to resolve possible naming conflicts (homonyms or synonyms) and specify equivalencies and overlaps between classes and attributes. The case of overlap between categories is resolved by splitting them into disjoint classes. Their common part forms a new class. For example, Fig.1 shows the case of decomposing two overlapping classes: “Industrial, commercial and transport units” (CORINE Land Cover) and “Technical and transport infrastructures” (CLUSTERS, 1995) into three disjoint classes: “Industrial or commercial units”, “Transport infrastructures” and “Technical infrastructures”. In this way, semantic factoring decomposes complex concepts into the simpler concepts out of which they are constructed.

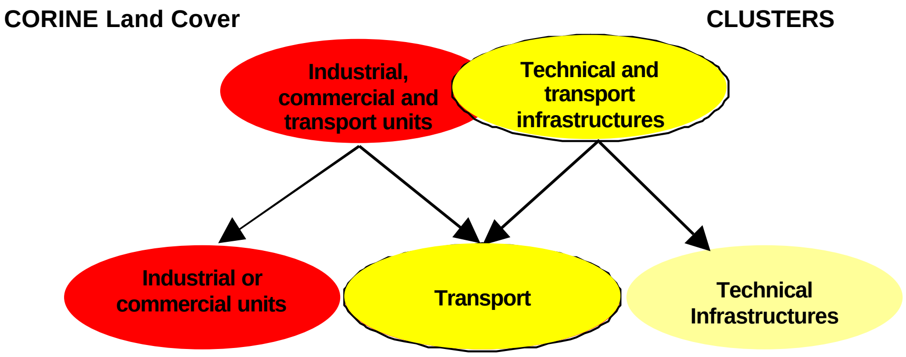

CORINE Land Cover
CLUSTERS
Industrial, commercial and transport units
Technical and transport infrastructures
Industrial or commercial units
Transport
Technical Infrastructures

Fig. 1. Semantic factoring of overlapping classes

At the second step, the basic notions and algorithms of *Formal Concept Analysis* (Wille, 1992; Ganter & Wille, 1999) are used, in order to combine the fundamental categories derived by the process of Semantic Factoring, as well as their properties, and generate what is called a Concept Lattice. The basic concepts of Formal Concept Analysis are:

- A *Formal Context* (G, M, I) is a set of objects G, a set of attributes M and a binary incidence relation I.
- An *Incidence Relation* I, or *gIm* is the connection between objects and attributes
- A *Formal Concept*, *Conceptual Class* or *Category* is a collection of entities or objects exhibiting one or more common properties or characteristics:
- A *Superconcept/subconcept relation* is the order proceeding top-down from more generalized concepts to more specialized concepts:
  $(A_1, B_1) \leq (A_2, B_2)$ if $A_1 \subseteq A_2$.
- A *Concept Lattice* {B (G, M, I); ≤} is the ordered set of all formal concepts of a formal context.

Concept Lattices are used to formalize geospatial concepts and relationships and generate a single integrated structure from different categorizations, in order to reveal their association and interaction. Concept Lattices are rich structures, since they allow the existence of overlapping relationships between formal concepts. Besides the original categories, concept lattices include additional ones, which result from the decomposition or fusion of original categories and make it more symmetric. Furthermore, Formal Concept Analysis helps to detect possible implications between final classes, which are not pre-defined, as well as to reveal hierarchical relationships, which were not initially obvious. The integrated Concept Lattice of the running example is shown in Fig. 2.

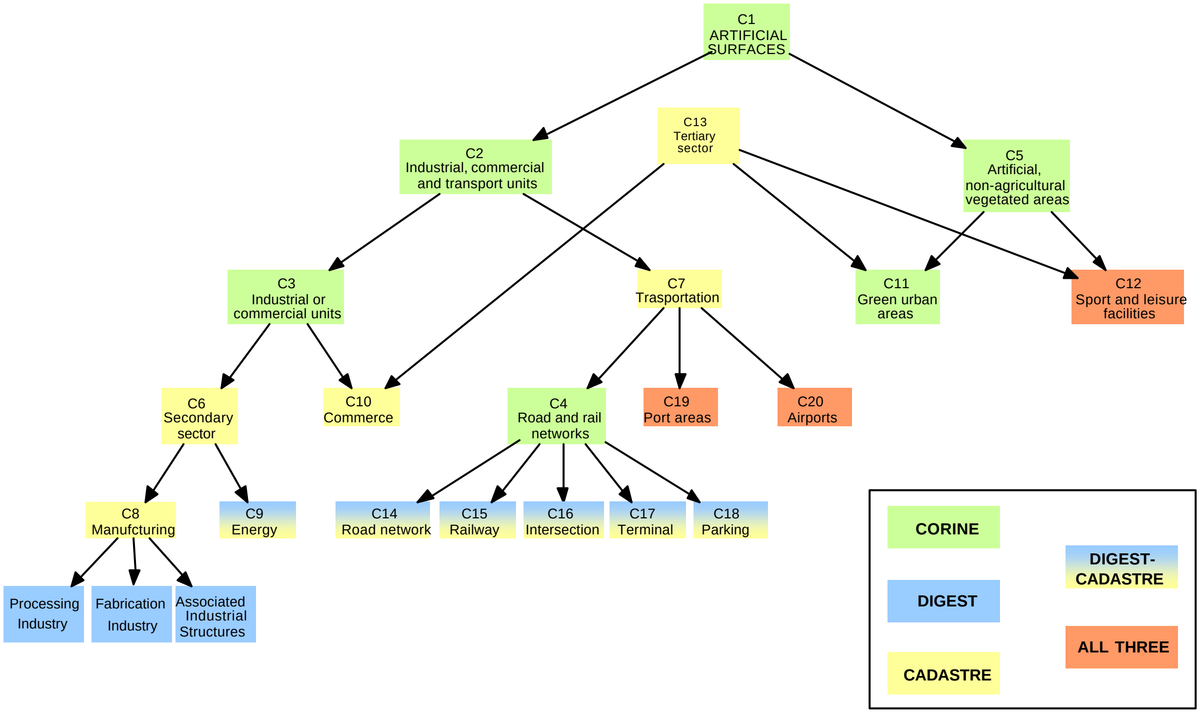


Fig. 2. Integrated Concept Lattice

Development of a Multi-Scale, Multi-Context Database

The integration of different, existing categorizations provides a flexible and effective means to build a multi-scale, multi-context database (Kokla & Kavouras, 1999). The integration can proceed both to the "vertical" and the "horizontal" direction (Fig. 3), which refer to different levels of detail (vertical integration) and different application contexts (horizontal integration). Thus, the integration methodology can be used in model generalization, in order to provide the means to move along different levels of detail and intelligently change scale, but also to move across different contexts and perform a change in the perception of geographic information.

The integrated Concept Lattice links similar classes of different levels of detail, and thus serves as a guide for the determination of the appropriate schema for a specific scale and context through interpolation (Fig. 4).

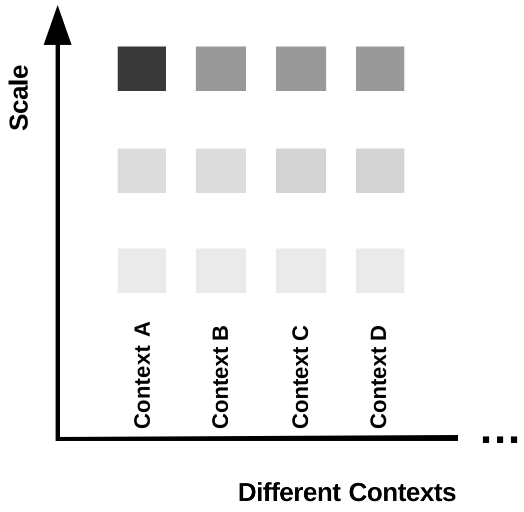


Fig. 3. Integration along different scales and across different contexts

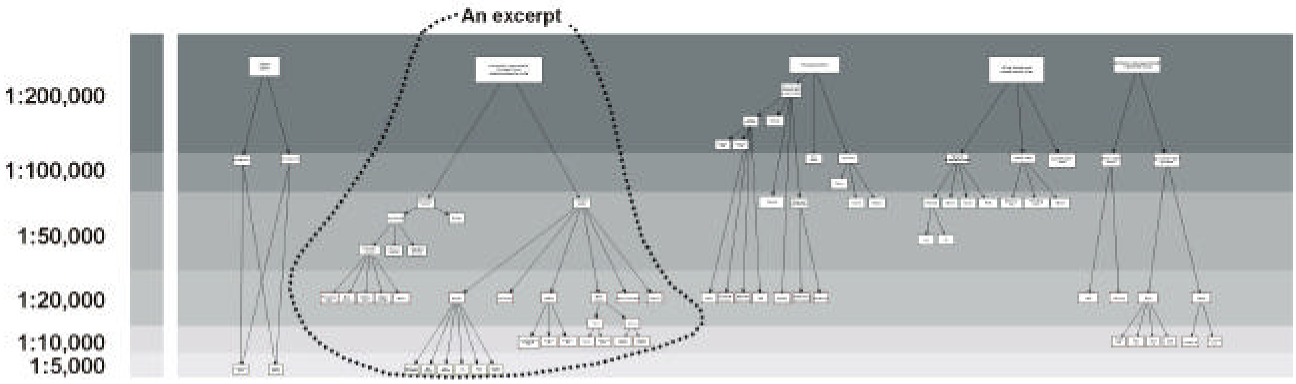
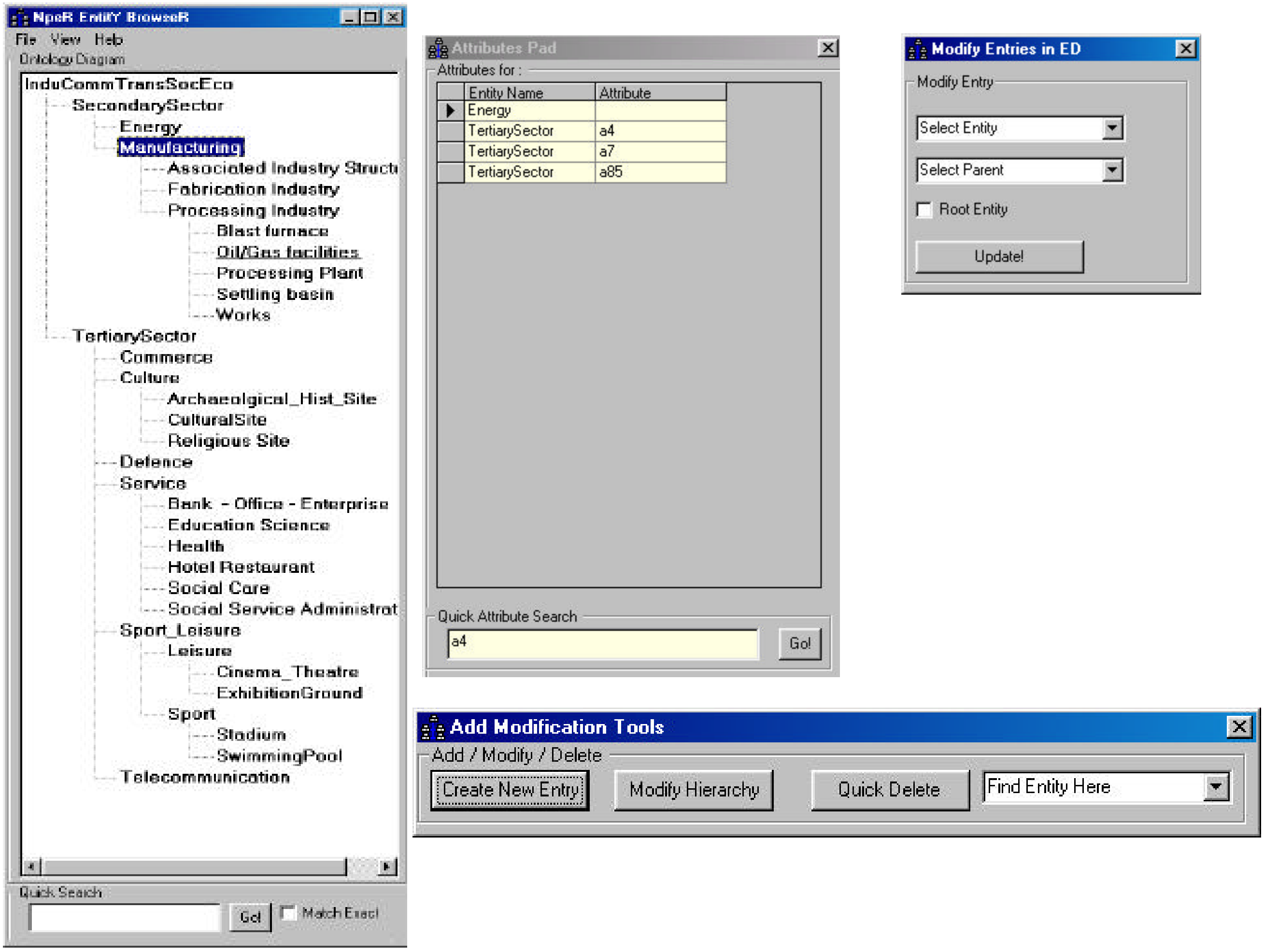


Fig. 4. The final multi-scale, multi-context database

Ontology Browser

Moreover, in order to be able to create and manipulate geographic ontologies, an Ontology Browser (Perdikis, 2000) has been developed as part of the research. The software has the ability to load and save geographic ontologies, create, modify, delete categories and attributes, search for specific categories according to their name or any of their attributes and trace the full hierarchy of a category (Fig. 5).

The software is build in the MS Windows environment, using an SQL database to store,



- InduCommTransSocEco
  - SecondarySector
    - Energy
    - Manufacturing
      - Associated Industry Structu
      - Fabrication Industry
      - Processing Industry
        - Blast furnace
        - Oil/Gas facilities
        - Processing Plant
        - Settling basin
        - Works
  - TertiarySector
    - Commerce
    - Culture
      - Archaeolgical_Hist_Site
      - CulturalSite
      - Religious Site
    - Defence
    - Service
      - Bank - Office - Enterprise
      - Education Science
      - Health
      - Hotel Restaurant
      - Social Care
      - Social Service Administrat
    - Sport_Leisure
      - Leisure
        - Cinema_Theatre
        - ExhibitionGround
      - Sport
        - Stadium
        - SwimmingPool
    - Telecommunication

	Entity Name	Attribute
▶	Energy	
	TertiarySector	a4
	TertiarySector	a7
	TertiarySector	a85

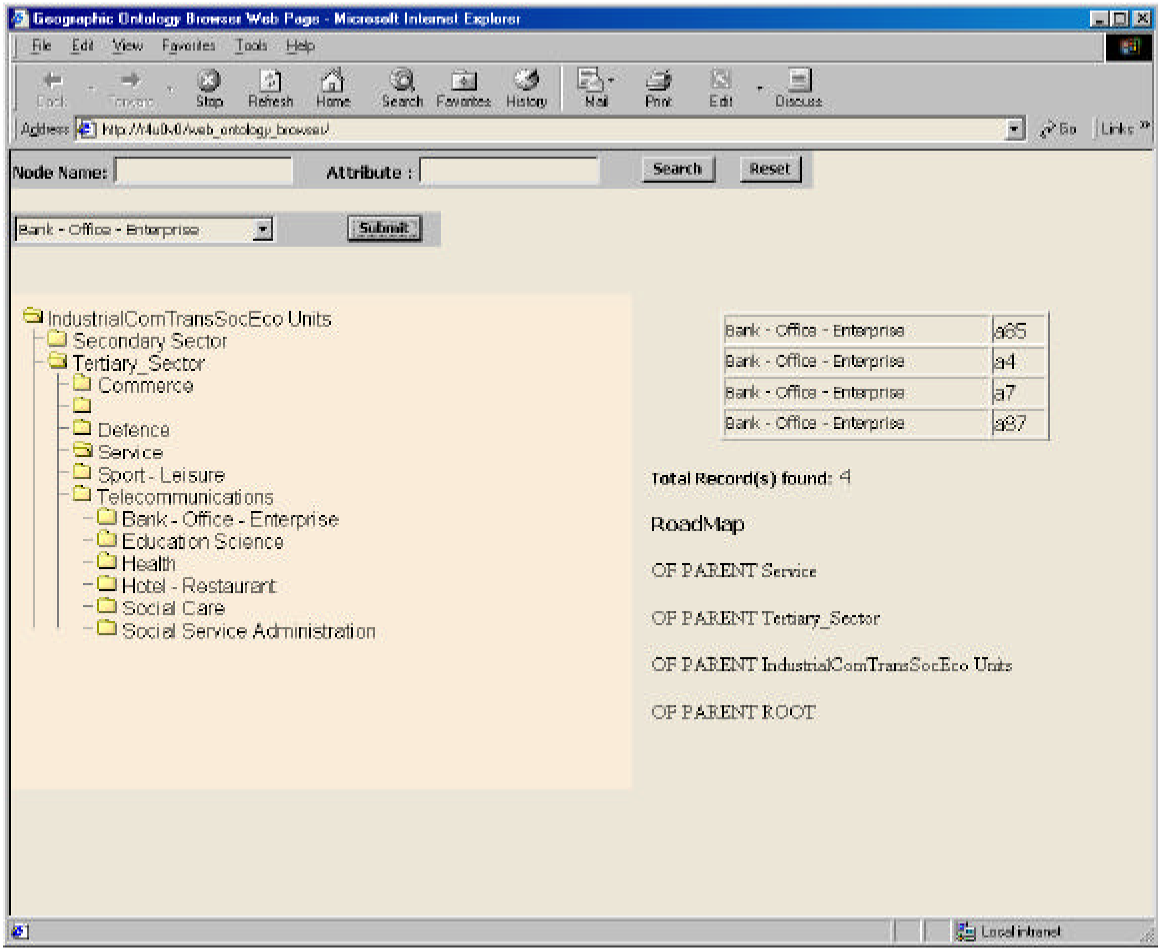
Quick Attribute Search: a4

Select Entity
Select Parent
□ Root Entity

Find Entity Here

Fig. 5. Ontology Browser

retrieve and query the data, a Graphic Windows Control (OCX) to display ontologies and all the graphic tools to support its functionality. The topology of the ontologies (i.e., inheritance of attributes and parent identification) is selectively build real-time or explicitly through menu commands of the software.
Moreover, a Web Interface (Fig. 6) has also been developed with the ability to display and query ontologies. This consists of Active Server Pages that interface the database of the software and display the ontologies in real-time using a Java applet, as the ontologies are build or modified.


Geographic Ontology Browser Web Page - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss
Address http://r4u0v0/web_ontology_browser/
Go Links
Search Reset Submit
Local intranet

Node Name:

Attribute :

Bank - Office - Enterprise

- IndustrialComTransSocEco Units
  - Secondary Sector
  - Tertiary_Sector
    - Commerce
    -
    - Defence
    - Service
    - Sport - Leisure
    - Telecommunications
      - Bank - Office - Enterprise
      - Education Science
      - Health
      - Hotel - Restaurant
      - Social Care
      - Social Service Administration

Bank - Office - Enterprise	a65
Bank - Office - Enterprise	a4
Bank - Office - Enterprise	a7
Bank - Office - Enterprise	a87

Total Record(s) found: 4

RoadMap

OF PARENT Service

OF PARENT Tertiary_Sector

OF PARENT IndustrialComTransSocEco Units

OF PARENT ROOT

Fig. 6. The Web Interface

EVALUATION

The proposed methodology provides a suitable tool for information formalization, integration and generalization. More explicitly, the integrated Concept Lattice is not strictly tree-structured, since certain classes may have more than one superclass. This flexibility of the integrated Concept Lattice permits its use for different applications. This means that hierarchies are used as conceptual tool and not as restriction of the methodology. For example, class «Commerce» (C10 in Fig. 2) may at different circumstances belong, either to «Industrial or commercial units», or to «Tertiary sector».
Furthermore, the methodology can be successfully applied independently of the spatial and thematic resolution represented by the input classification schemata. Therefore, it is possible to associate classifications created for similar purposes dealing with many overlaps between the input classes or, to integrate classification schemata of different thematic resolutions.
Moreover, the method helps to identify and resolve heterogeneities between original categorizations. These refer to schematic heterogeneities due to different structures of the

original generalization hierarchies, or due to definition of similar classes at different levels of detail, and to semantic heterogeneities caused by overlapping definitions of similar classes.

Finally, the integration process converts the input classification schemata to a single schema corresponding to an integrated, but also uncompromising conception of space. Namely, the original classes and attributes are not altered, but semantically related to each other to form the final hierarchical schema. Therefore, the integration process identifies similarities and reconciles differences without preventing the independent and autonomous use of the original schemata.

REFERENCES

BISHR Y., "Overcoming the semantic and other barriers to GIS interoperability," *International Journal of Geographical Information Science, P. Fisher, M. Armstrong and B. Lees (ed.)*, Special Issue: Interoperability in GIS, edited by. A. Vckovski, Vol. 12, No 4, 299-314, Taylor & Francis, June 1998.

CORINE Land Cover-Technical Guide, Published by the European Commission, EUR 12585 EN, Luxembourg, 1994.

CLUSTERS: Classification for Land Uses STatistics: EUROSTAT's Remote Sensing programme", 1995 http://europa.eu.int/comm/dg06/publi/landscape/tab2_5.htm

DIGITAL GEOGRAPHIC INFORMATION WORKING GROUP (DGIWG). "Digital Geographic Information Exchange Standard (DIGEST) Standards Specification", Part 4, Edition 2.0, NIMA, June 1997.

GANTER B., and WILLE R., *Formal Concept Analysis, Mathematical Foundations*, Springer-Verlag, Berlin Heidelberg, 1999.

GUARINO N., "Formal Ontology in Information Systems," *in Formal Ontology in Information Systems, Proc. of the 1st Int. Conf. (FOIS'98)*, N. Guarino (ed.), June 6-8, 1998, Trento, Italy.

HELLENIC MAPPING AND CADASTRAL ORGANIZATION. Ministry of the Environment, Planning and Planning Works. "Technical Specifications for the Greek National Cadastre-Land Use Classification," 1996 (in Greek).

KOKLA M., & KAVOURAS M., "Spatial Concept Lattices: An Integration Method in Model Generalization," *Cartographic Perspectives*, North American Cartographic Information Society, Number 34, Fall 1999.

PERDIKIS N. "Ontology Browser", MSc. Dissertation, National Technical University of Athens, Department of Rural and Surveying Engineering, 2000 (in Greek).

SMITH B., "Basic Concepts of Formal Ontology," *in Formal Ontology in Information Systems, Proc. of the 1st Int. Conf. (FOIS'98)*, N. Guarino (ed.), June 6-8, 1998, Trento, Italy.

SOWA, J. F., *Knowledge Representation: Logical, Philosophical and Computational Foundations*, Brooks/Cole, USA, 2000.

WILLE, R. "Concept Lattices and Conceptual Knowledge Systems," *Computers and Mathematics with Applications*, Vol.23-6-9: 493-515, 1992.