COGNITIVE AND CREATIVE FRAMEWORK FOR DIGITAL EARTH

A.A. Liouty
Institute of Geography of the Russian Academy of Sciences,
29, Staromonetny per., Moscow, 109017, Russia, Fax: (7095) 959-00-33
A.I. Martynenko
Institute for Informatics Problems of the Russian Academy of Sciences,
2 build., 44,Vavilova st., Moscow, 117333, Russia, Fax**:** (7095) 930-45-05,
E-mail: a_martynenko@mail.ru
I.M. Zatsman
Institute for Informatics Problems of the Russian Academy of Sciences,
2 build., 44, Vavilova st., Moscow, 117333, Russia, Fax**:** (7095) 930-45-05,
E-mail: zatsman@bur.oivta.ru

Abstract

The Global Geoinformatic Mapping appears as the prior direction of integration of georeferenced data. Its goal is to represent of the world and create of the global digital model of the Earth (denoted as Digital Earth) comprised of million of multidimensional space images and electronic topographical and thematic maps. The need in collecting more complete and precise georeferenced data about the Earth phenomena and objects highly increases today. As a result the new information technologies and techniques are required to search and retrieve of georeferenced data. Search and retrieval of georeferenced data will appear as be semantics-sensitive procedures in Digital Earth.

1 Introduction

Today, at the boundary of millennia, the Global Geoinformatic Mapping (GGM) appears as the prior direction of scientific and technical progress. Its goal is the cartographic representation of the real world and creation of the global computer model of the Earth, comprised of millions of space images and electronic maps of various subjects and scales, themes and reference information. This fundamental problem can be solved by cartographers from various countries, who should meet the 21 century as partners, possessing new ideas, courage and intellectual technologies for creation and application of maps.

The need in collecting more complete and precise data about Earth phenomena and objects highly increases today. As a result of development of space survey, the bulk of geospatial information grows very quickly. So the new electronic technologies and standards are required to process spatial data and deliver it to the users. This means the increasing role of the intelligence. GGM is aimed at preparation and fulfilment of measures to develop and implement conceptual and methodological basis, normative and legislative documents and standards of metadata for geographic, geodetic, gravimetric, space, photogrammetric and cartographic information, electronic photomaps and spatial (3D) terrain models, formats for spatial data interchange; to develop and implement methods, hard- and software tools and technologies for acquisition, storage, analysis and processing digital cartographic data, creation of conventional (paper) and electronic maps; to develop and implement Base of metadata and Bank of spatial data, digital and electronic maps, geoinformation systems of various applications. At the modern stage, GGM is closely connected with the development of geography, geodesy, remote sensing, photogrammetry and cartography.

Since 1992 the Electronic Map System to cover all the Earth's surface has been developed in Russia. The System can be considered a digital library of maps integrated on a single concept, coordinated scales, projections, and coordinate system as well as content and the legend. The library is being created on the

basis of available maps, space images and other information sources about the Earth in accordance with demands of various users. The Electronic Map System is based on the following principles [1, 2]:

- system approach as the conceptual basis for creation and implementation of cartographic models, as the methodology of research and projecting the System, and as the scientific method of development of effective computer-aided multimedia technologies;
- principle of mathematical and cartographic modeling as a way to visualize terrain features and objects;
- principle of raster input/output of cartographic data, and vector data processing;
- principle of controllability of digital cartographic data;
- principle of utmost complete acquisition, one-time exhaustive analytic/synthetic spatial data processing and its use by many users.

To implement Digital Earth, the following tasks must be solved:

- investigation and generalization of the international experience in Earth mapping and application of geospatial data;
- development of cartographic thinking, revelation and exploration of objective regularities and features of Earth mapping, formulation of principles of GGM on this base;
- investigation and elaboration of the concepts, intelligent methods and technologies of acquisition, integration, analysis, processing, modeling, displaying of electronic maps, 3D-images, dynamic cartographic models, virtual maps;
- development of basic categories and notions, international and national standards in the area of GGM;
- creation of cognitive and creative framework for Digital Earth, the International Earth Knowledge Management System, Global Base of Metadata and Banks of Geospatial Data, telecommunication networks;
- development and implementation of high-skilled cartographic expert systems and GIS of various intent;
- increasing professional skill level of the personnel in the area of GIS and intellectual technologies at all educational levels;
- revelation of contradictions in the area of GGM, determination of directions of its evolution, elaboration of the criteria and methods for evaluation of the effectiveness of the new technologies;
- development and implementation of the international cooperation programs in the area of GGM strategies including participation of governmental and private cartographic enterprises of countries - members of the ICA.

The paper discusses cognitive and creative framework for Digital Earth (DE). The choice of a cognitive basis will determine the methods of knowledge representation and semantic search and retrieval in DE. The main difference between the proposed cognitive basis from the design concept of the Electronic Map System is the use of three lingual systems: verbal, geolanguage and formal lingual systems. It is assumed that knowledge in the Earth sciences can be adequately represented with the help of the combination of three lingual systems. An integration of semiotic spaces of these lingual systems can be referred to as the semiotic sphere which content is described in the present paper.

The paper is aimed at structurization of the semiotic sphere into semiotic spaces and analysis of the relationship of the map language with the lingual systems of DE. The semiotic approach underlies the analysis of the following items to create cognitive and creative framework for DE which are directly relevant to knowledge representation in the form of signs:

- semiotic heterogeneity and translation of non-verbal texts of DE;
- declarative-lingual and procedural approaches to geodata search and retrieval;
- semiotic problems of DE;
- structurization of the semiotic sphere.

2 Semiotic heterogeneity and geotext translation

In [3] the generalized definition of sign is suggested as coming from the necessity of sign representations of full-text scientific documents in the process of their logic and semantic modeling. Semiotic aspects of the creation of DE are approached from the generalized concept of sign. In the present paper, the generalized concept of sign is used to describe sign representations of electronic forms of discrete and continuous geoinformation.

The traditional encoding of geoinformation in digital raster and/or vector forms is limited in relation to semantic search and retrieval of geodata. It can be explained by the fact that the search of heterogeneous verbal and non-verbal information is mainly determined by the employed languages and semiotic systems. The essence of the cognitive and creative framework for DE is as follows: DE is viewed as a digital library including geotexts in different geolanguages. Besides, DE comprises verbal information positioned in time and space, i.e., georeferenced verbal texts in various natural languages as well as texts in formal languages.

Verbal languages have their grammar descriptions. The system of paradigmatic, syntagmatic and semantic relations has been studied and described for many centuries. This knowledge is made use of in verbal thesauri and semantic search and retrieval of verbal information. However, in geolanguages, the situation is completely different.

There are some types of geotexts which can be adequately translated into the natural language and vice versa. For example, the translation of a verbal text into the graphical language is used to describe lithological stratigraphical sections [4, pp.46-48]. The translation is a part of the description of a geological phenomenon referred to as zhaltauskaya suite. The suite description consists of a non-verbal text, geological schemes and a series of lithological stratigraphical sections. Information on single sections is first given in a verbal form followed by the translation into a graphical form. The translation into a graphical language is possible due to the linearity of the geolanguage of lithological stratigraphical sections.

In a general case, for geotexts, the degree of translation and its adequacy for different languages can differ dramatically. For a large number of geotexts the possibility of an adequate translation is missing. For example, in Earth sciences, widely used geological maps and schemes referred to as geotexts in the map language are next to impossible to translate into verbal languages. In contrast to lithological stratigraphical sections and their comparison schemes, such geotexts as geological schemes and maps are difficult to translate into a natural language due to the following common features of texts in the map language [5]:

- graphical statements as combinations of graphical signs can be discrete and continuous;
- linear organization in sign combinations which are not always can be separated and determined is missing.

These features are special values of the following semiotic characteristics of electronic forms of graphical texts considered in [3]:

1) certain/uncertain forms of signs,
2) statical/dynamical signs,
3) one-dimensional/two-dimensional/multidimensional signs,
4) linear/ordered/unordered sign combinations,
5) discrete/continuous signs and sign combinations.

It follows then that, in a general case, sign representations of electronic forms of geotexts make use of certain, fuzzy, random, uncertain, statical, dynamical, one-dimensional, two-dimensional, and multidimensional sign forms. Sign combinations can be linear, ordered, unordered, discrete, continuous or discrete-continuous. Therefore, the basic conclusion is as follows: (1) geotexts can be translated into verbal languages in case of specific values of their semiotic features. See the example of lithological

stratigraphical sections and their comparison schemes. It should be stressed that (2) semantic aspects of geotexts cannot be always defined by metadata.

The last two statements concerning translation and metadata are specific cases of the key semiotic statement on the three basic spheres of knowledge representation [6]:
1) non-verbal knowledge which cannot be represented in a linguistic form,
2) verbal knowledge which cannot be adequately translated into a non-verbal form,
3) knowledge which can be represented in both verbal and non-verbal forms.

The first case when the adequate translation into verbal languages is not possible and semantics of geotexts is not represented in metadata seems to be the most complicated one in semantic geotext modelling and information search and retrieval in DE.

3 Declarative Lingual and Procedural Approaches to Geodata Search

In terms of semantic search of geodata concepts of DE can be divided into two categories:
- procedural framework of search;
- declarative lingual framework of search.

The first approach can be illustrated by the concept of Geobrain. The concept is based on the object-oriented method, the essence of the method being the integration of semantic description of geoinformation in DE represented as geoobjects. The geoobject is defined as a combination of initial geodata, e.g. a geological map or a scheme, their attributes, e.g. metadata of the map, as well as a set of processing methods of initial geodata. There are two levels of storing and searching geoinformation: at the first level only metadata are stored while at the second level geoobjects are stored (the level of the distributed archive of geoobjects which are connected with each other by some relationship) [7].

The search concept of the first framework mainly makes use of the procedural approach to semantic search of initial geodata which cannot be adequately translated into verbal languages and which semantics is not expressed in metadata. The projects in this field will depend upon processing methods of initial geodata to be used in projects.

If initial geodata can be presented as geotexts in geolanguages, the declarative lingual approach to DE and semantic search and retrieval is employed. In the system of geolanguages of DE the central part belongs to the map language. Currently, the studies of the map language mechanism, structure and functional principles are at the initial stage [8]. Other languages referred to as a wholesome semiotic system are not studied at all.

The spectrum of geodata which can be considered as geotexts is rather wide starting from lithological stratigraphical cross-sections to topographical and problem-oriented maps and schemes. Is it possible to introduce a generalized concept of sign which will allow one to interpret any initial geodata in DE as geotexts? In classical semiotics a positive answer can be given in case of a relatively stable system of relationship and/or models being a framework for defining the semantics of sign based on its form [9]. A possible variant of a stable system is a verbal&visual thesaurus [10].

The fundamentals of the search and retrieval strategy based on the declarative lingual approach are discussed in [10, 11] and the principle of multilevel semantic modelling and sign representation of texts are proposed. Let us enumerate the basic system principles to be used as a basis for multilevel semantic models (MSM) in DE:
1) Geotext MSM integrate the map language as a system with a grammar structure and a hierarchy of lingual elements.
2) In MSM, electronic maps and other coordinated geoimages of various subject domains are considered basic texts while space photos, air shots and other non-map geoimages are referred to as supplementary geotexts. Non-map geoimages can be correlated in space and time with basic geotexts with scales being differentiated.

3) The tasks of modelling, indexing and initial geodata search and retrieval are solved for basic geotexts. Supplementary geotexts are made use of at the final stage when found data are visualized.

4) In the process of multilevel semantic modelling the notions of an object to designate the physical nature of the phenomenon, a concept and a generalized sign to represent semantics in DE are set apart. In [3], to designate generalized signs, the term "sign-sets" was proposed.

5) For a general case, within MSM, user's information needs to search necessary initial data on geoobjects are expressed, as a user's combination of sign-sets. The user's combination may not have the adequate sign-set combination in DE. In this situation, when the search query with a specified correspondence degree is executed, the user gets a zero result.

The multilevel semantic model architecture features a classification of geoobjects, dictionaries of graph forms, a geothesaurus and metadata catalogues.

4 Semiotic problems of DE construction

The next stage of the declarative lingual approach is aimed at the description of the semiotic sphere of DE and development of a geothesaurus. The geothesaurus of DE is expected to be used to search and retrieve geodata and to solve the following semiotic problems:

- to normalize verbal and graphical signs used for semantic encoding and indexing verbal texts and geotexts (normalization problem),
- to express explicitly the lingual belonging of signs in geolanguages (explicitness problem),
- to define in context the values of proprietary and semidefinite signs which are not universally accepted but nevertheless used in many geotexts (context definition problem).

Sign representations of geotexts are proposed to be obtained on the basis of normalized sign combinations, which are descriptors of the geothesaurus, with the help of the semiotic approximation method [10]. Thus, the geothesaurus is a source of normalized graphical signs and its descriptors are used for sign representation of geotexts comprised in DE.

In addition to sign normalization the geothesaurus can be used to define the values of proprietary signs, with geotext context being used in the following way. To define the values of proprietary signs of a geotext, a correspondence between the forms and/or sign images and descriptors in the geothesaurus which are close in semantics can be established. For semidefinite signs another approach should be applied. The values of proprietary signs are normally defined in the text while the values of semidefinite signs are sometimes not disclosed. Thus suggests taking into account the special features of semidefinite signs when creating the semiotic sphere of DE. Therefore, DE should be supplied with materials to describe semidefinite signs [10].

It is possible, however, that the geothesaurus won't feature the corresponding descriptors to create sign representations of geotexts and/or to define the values of proprietary and semidefinite signs. When geotexts with proprietary signs are processed the geotext semantics can be used to expand the geothesaurus through necessary descriptors. When geotexts with semidefinite signs are processed supplementary materials describing their values can be used. Uncertain boundaries between geolanguages can be explicitly expressed by multiple correspondence of descriptors and several languages or lingual systems. This allows one to obtain explicit presentation of sign lingual belonging in the system of geolanguages.

5 Structurization of the semiotic sphere of DE

In the beginning of the paper the semiotic sphere was defined as an integration of semiotic spaces of the following lingual systems:

- verbal lingual system,
- geolanguage system,
- formal lingual system.

The above-mentioned paragraphs of the paper allows one to expand the definition by specifying "integration of semiotic spaces". In accordance with the theory of the semiotic sphere developed by Yu.

M. Lotman, the texts corresponding to semiotic spaces can be intertranslated, partially translated and nontranslated [12].

Text translation both within the geolanguage system and between the verbal and geolanguage systems is one of the features of semiotic space integration. The research of the intertranslation problem is at the initial stage. However, it has been established that some types of geotexts, e.g. lithological cross-sections, are first described verbally and then transformed into a graphic form. Therefore, in the semiotic sphere of DE, verbal texts are sometimes adequately translated into geotexts and vice versa. Verbal texts and maps cannot be intertranslated. The second feature of the integration is as follows: texts with uncertain lingual belonging stored in DE together with texts with a clear verbal, geolanguage or formal lingual belonging of signs. The third feature of integration implies that the map language consists of three sublanguages specifying and/or expressing [8]:

- spatio-temporal characteristics of objects and phenomena through a formal lingual system,
- content of objects and phenomena expressed through the sign system of "map symbolism",
- place names and terms normally represented by the natural language signs.

Therefore, the semiotic space of the map language intersects with the semiotic spaces of verbal and formal lingual systems. Taking into account the three features the semiotic sphere of DE can be defined as an integration of semiotic partially intersected spaces of the three lingual systems with the map language being the core. The semiotic sphere of DE can be metaphorically illustrated as three partially intersected circles. The semiotic space of the map language is at the intersection (Fig. 1).
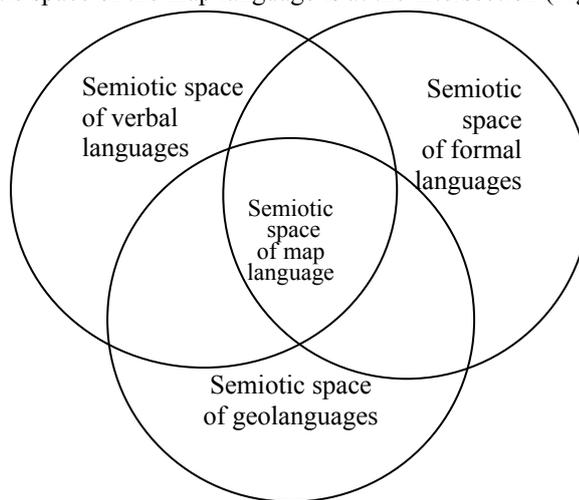

Fig. 1. Semiotic spaces of DE.

The sign systems for electronic maps and other geotexts are based on the sign model of the semiotic sphere. Thus, the choice of the sign model is vital to create the cognitive and creative framework for DE. The declarative lingual approach suggests building a sign model taking into account all the characteristics: certain/uncertain forms of signs, statical/dynamical signs, one-dimensional/two-dimensional/multidimensional signs, linear/ordered/unordered sign combinations, discrete/continuous signs and sign combinations. This list is characterized by the first semiotic feature which can take the following values of sign forms: certain, multivariant, random, and fuzzy [11, 13].

M.F. Goodchild provides an introduction to the problem of uncertainty in geographic data. More general frameworks for uncertainty were needed, including fuzzy sets, Bayesian statistics, subjective probability [14]. The list of concept problems associated with DE proposed by Goodchild [15] includes the following statement: "The role of vague information in ED needs to be resolved...". Due to the inclusion of certainty/uncertainty characteristics into the sign model the information capacity of signs in DE has dramatically increased. In this model uncertainty is expressed in declarative lingual form through

descriptors of the verbal&visual geothesaurus. The coordination of methods to build the geothesaurus can become the basis for an international effort to begin construction of cognitive and creative framework for DE and to shape common vision of DE.

6 Discussion

Semiotic features proposed to create a new model of the semiotic sphere of DE are characterized by the abstract concept of "sign-set". The concept of sign-set will be used to cover the above-mentioned spectrum of sign form characteristics. In terms of mathematics, sign-sets can be analyzed with the help of the set theory and formal logic procedures can be applied. In terms of semiotics, sign-set can be referred to as non-verbal signs and their contents can be used for knowledge representation in DE. Sign-sets with multidimensional forms are very difficult to include into traditional paper articles. For example, in digital climatic maps, dynamic, three-dimensional and fuzzy sign-sets might be convenient to represent continuous change in cloudiness for some period. However, in traditional paper editions, this is next to impossible. To designate sign-sets in traditional paper documents, the corresponding metasigns can be used in the form of indexed letters of the alphabet. In computer information processing, metasigns can be also used as indicators of the place where sign-sets are stored. The relationship between the metasign, sign-set, referent, and concept can be schematically reproduced as a tetrahedron (Fig. 2) and is referred to as a generalization of the semiotic triangle by Frege [8].

Fig. 2. Semiotic tetrahedron
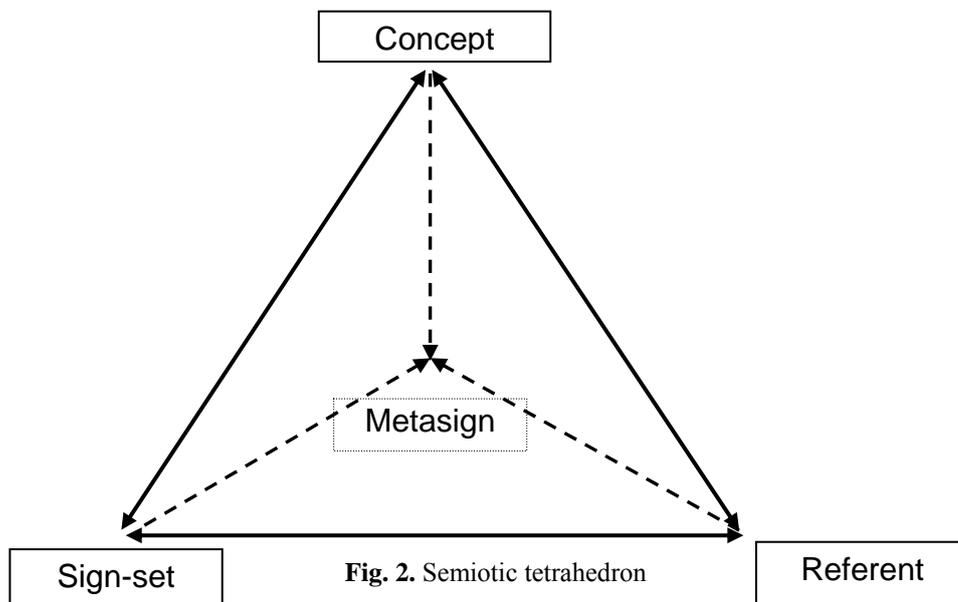
The presentation of geotexts in declarative lingual form with the use of sign-sets will provide for the expansion of semantic space to search geodata in DE. In theory there is a possibility to represent uncertain information of various types in the sign form. However, computer implementation of this approach calls for extra R&D.

7 Conclusions

There are techniques and software to search and retrieve of geotext via metadata. In the techniques and software metadata are attributes describing the spatial descriptors and content characteristics of the geotext objects in a symbolic form. The attributes have specific semantic meanings and thus allow metadata semantic search. The main disadvantage of metadata approach is that the geoobjects have to be described in advance. To make a description, geotext objects have to be recognised in map and understood by the person doing the description. So the description results heavily depend on the person doing the meta-description.

The paper proposes declarative lingual approach to create cognitive and creative framework for DE and implement semantic search and retrieval of initial geodata. Essentially, the approach is as follows: DE is treated as a digital library comprising texts in the three lingual systems. The next stage of research is aimed at a new sign model to describe the semiotic sphere and provide foundations for verbal&visual geothesaurus of DE.

Acknowledgements

References

1 Martynenko A.I. The Concept of the Electronic Maps System // Proc. 16th International Cartographic Conference, Cologne, 1993, vol. 2, pp. 1160 – 1163.

2 Martynenko A.I. Global Geoinformatic Mapping as the Methodology and Technology of the Future // Proc. 19th International Cartographic Conference, Ottawa, 1999, vol. 1, pp. 39 - 42.

3 Zatsman I.M. Semiotic Problems of Modeling and Search of Full-Text Scientific Documents//Trudy Mezhdunarodnogo Seminara DIALOG 2001 po Kompyutornoi Lingvistike i Ee Prilozheniyam (Proc. of the International Workshop on Computer Linguistics and Applications DIALOGUE - 2001).- Moscow, 2001.

4 Zaitsev Yu.A., Kheraskova T.N. Vend Tsentralnogo Khasakstana (The Vendian of Central Khazakstan)- Moscow, MGU, 1979.

5 Schlichtmann H. Specific traits of the sign system "map symbolism" // Paper at 11th International Cartographic Conference. Warsaw: July 29 to August 4, 1982. 20p.

6 Eco U. A Theory of Semiotics. - Bloomington: Indiana University Press, 1976. - 356 pp.

7 Di L., McDonald K. Next Generation Data and Information Systems for Earth Sciences Research // Proc. International Symposium on Digital Earth. - Beijing: Science Press, 1999. - pp. 92-101.

8 Liouty A.A. Yazyk karty: sushchnost, sistema, funktsia (Map Language: the Essence, System, Function). - Moscow: IG RAN, 1988.

9 Solomonik A. Semiotika i Lingvistika (Semiotics and Linguistics). - Moscow, Molodaya gvardiya, 1995. - 352c.

10 Zatsman I.M. The Method of Semantic Approximation and Verbal&Visual Knowledge Representation in Digital Libraries // Sistemy i Sredstva Informatiki, Moscow, Nauka, 2001, N 11.

11 Zatsman I.M., Liouty A.A., Martynenko A.I. Semantic Search in Digital Geolibraries // Systemy i Sredstva Informatiki. - Moscow, Nauka, 2000, N 10. - pp. 193 - 205.

12 Lotman Y.M. Universe of the mind: a semiotic theory of culture. - London&New York: "I.B.Taurus&Co. Ltd", 1990. - 306 p.

13 Zatsman I.M. Semantic Encoding and Mark-Up of Geological and Geographical documents in Polithematic Digital Libraries//Informatsionnye Tehnologii, 2000, N 11. pp. 2-11.

14 Goodchild M.F. Introduction: special issue on "Uncertainty in geographic information systems" // Fuzzy Sets and Systems. - 2000. Vol. 113, N 1. - pp. 3-5.

15 Goodchild M.F. Implemententing Digital Earth: : A Research Agenda // Proc. International Symposium on Digital Earth. - Beijing: Science Press, 1999. - pp. 21-32.