THE PROBLEM OF TYPOLOGY OF SIGNS AND THE SEMANTIC SEARCH IN DIGITAL GEOLIBRARIES

Liouty A.A.,

Institute of Geography
of the Russian Academy
of Sciences,
29, Staromonetny per.,
Moscow, 109017, Russia
Fax: (7095) 959-00-33
E-mail: lut@map.msk.ru

Zatsman I.M.

Institute for Informatics Problems
of the Russian Academy
of Sciences,
30/6, Vavilova st., Moscow,
117900, Russia
Fax: (7095) 310-70-50,
E-mail: zatsman@mail.ru

Full-text documents on the Earth Sciences may include maps, charts, air- and spaceimages, map-type diagrams, etc. (denoted as non-verbal components), which can describe spatial entities, spatial-temporal flow of processes, and phenomena [1, 2]. In the documents, non-verbal components can be even more valuable than verbal ones. Non-verbal components of the documents represent portions of general semantic spaces of digital geolibraries. The problem of semantic search in digital geolibraries comprises the following tasks: decomposition and semantic representation of non-verbal components in digital forms; semantic mark-up and encoding of components; indexing of the documents and its components. To solve this problem, it is necessary to define a universal semiotic system for digital geolibrary documents [3]. A definition of the semiotic system is closely connected with a classification of sign modes. The proposed classification of sign modes takes into account five facets of spatial entities/phenomena, which are described in non-verbal components of the paper and computer-based documents:

1. The determinate/non-determined nature;
2. The discreteness/continuity;
3. The dimensionality;
4. The dynamics/statics;
5. The physical and chemical characteristics of entities/phenomena.

The first facet covers determinate, multivariant (for each variant a determinate algorithm may be specified), random, and fuzzy entities/phenomena. The second facet covers discrete, continuous, and discrete-continuous ones. The entities may be either unidimensional, two-dimensional or multidimensional (third facet). The fourth facet covers static and dynamic entities. The physical and chemical characteristics are sizes, mass, composition, etc. (for example, drainage size, frost mass, chemical group composition). This classification of sign modes covers almost all semiotic modalities for non-verbal components of paper and computer-based documents on the Earth Sciences.

The work is supported by the Russian Foundation for Basic Research, project № 99-05-65491.

References

1. Gardels K. A Comprehensive Data Model for Distributed, Heterogeneous Geographic Information (http://www.regis.berkeley.edu/gardels/geomodel_def.html).

2. Liouty A.A., Yazik karty: sushcnost, sistema, funktsia (Map Language: the Essence, System, Function). - Moscow: IGRAN, 1988.

3. Zatsman I.M. Semanticheskoe kodirovanie i razmetka geologo-geograficheskikh dokumentov v politematicheskikh elektronnykh bibliotekakh (Semantic coding and markup of documents on the Earth Sciences in polythematic digital libraries. // Informastionnye Technologii (Information Technologies). - 2000. - N 9 (in print).