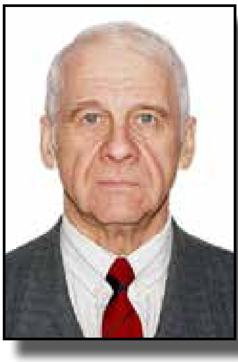
ORIGINAL ARTICLE
DOI: https://doi.org/10.30932/1992-3252-2024-22-2-1

World of Transport and Transportation, 2024,
Vol. 22, Iss. 2 (111), pp. 150–155

On Formation of Geospatial Knowledge Based on Transformation of Geo-Descriptions

Igor N. ROZENBERG



Sergey K. DULIN

Igor N. Rozenberg[1], Sergey K. Dulin[2]
[1] Russian University of Transport, Moscow, Russia.
[2] Federal Research Centre «Informatics and Control» of the Russian Academy of Sciences, Moscow, Russia.
[2] Joint Stock Company «Design & Research Institute for Information Technology, Signalling and Telecommunication on Railway Transport» (JSC «NIIAS»), Moscow, Russia.
✉ *[2] skdulin@mail.ru.*

ABSTRACT

An approach presented within the framework of the problem of formalisation and presentation of geoinformation structures in the GIS environment and based on the functional programming capacity is intended to simplify the techniques of dealing with characteristics of geodata attributes. The paper considers issues related to formation, comparison and transformation of geo-descriptions for updating geoinformation in the GIS environment. To implement the transformation of geo-descriptions, a simplification of formation of a tool for representing geoinformation and geo-text sets is proposed. The approach described in the paper is based on the choice of a functional programming paradigm implemented in the form of data analytics machine, which allows implementing transformations of geo-descriptions in various GIS languages and maintaining the geodata repository in an up-to-date state.

Keywords: GIS; formation, comparison and transformation of geo-descriptions; functional programming tools.

For citation: Rozenberg, I. N., Dulin, S. K. On Formation of Geospatial Knowledge Based on Transformation of Geo-Descriptions. World of Transport and Transportation, 2024, Vol. 22, Iss. 2 (111), pp. 150–155. DOI: https://doi.org/10.30932/1992-3252-2024-22-2-1.

The original text of the article in Russian is published in the first part of the issue.
Текст статьи на русском языке публикуется в первой части данного выпуска.

BACKGROUND

Processing of geoinformation applying a symbology layer causes constraints in implementation of semantic processing of GIS resources due to weak standardisation of attributive and visual geoinformation and poor-quality structures of its description, which are necessary for creation of an interconnected set of GIS resources. Descriptions of geodata, due to their heterogeneity, cannot be presented exclusively in linguistic form, therefore, the idea of developing an approach based on functional programming may be productive, which would allow to match geodata with encapsulated functions of their processing, that to some extent reflect their semantics [1; 2]. The proposed approach will allow to expand the system of used geo-descriptions pertaining to various GIS languages, including the presentation of visual geospatial knowledge.

RESULTS

Two Scientific Directions in Development of Cartographic Systems

The synthesis of cartography and semiotics gave rise to a new scientific discipline, i. e. cartosemiotics, which included the analysis, interpretation and transformation of symbolic cartographic systems.

Cartosemiotics uses linguistic descriptions in geoinformatics and makes it possible to represent attributive characteristics of maps, such as symbols and legends, using special linguistic construction of maps with model and content components.

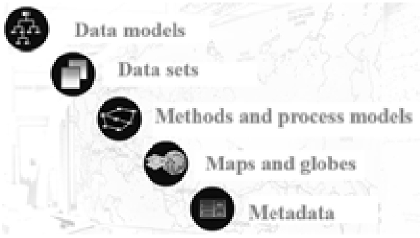
Over time, five sections were recorded in cartosemiotics, one way or another related to creation of geo-descriptions: cartosemiotic, cartolinguistic, formal-logical, cybernetic and sublinguistic. However, the formal-logical and sublinguistic sections have not undergone significant development and are marginal in nature.

Cartosemiotics, being essentially applied semiotics, includes three components similar to semiotics: 1) cartographic syntax of system of signs; 2) cartographic semantics associated with the analysis of meaning of the signs used, corresponding to the described GIS resources; 3) cartographic pragmatics, which allows us to measure the utility of the signs used depending on the level of knowledge and professional background of users of GIS resources. The central element in cartosemiotics refers to development of a map language – a correct and consistent symbology system designed to display cartographic information. The quality of the map language is characterised by the level of development of the means of presenting geospatial knowledge and by the possibility of visualising cartographic information. The most significant contribution to development of the map language was made by A. A. Lyuty [3]. The map language he created includes two sublanguages: one allows us to present the composition of geospatial objects (geo-bojecdts), and the other represents the geometry and topology of geospatial objects. Increased requirements for management of geospatial resources have led to a rapid growth in proposals and additions in the field of developing geospatial standards, which have made it possible to more fully and accurately form representations of geospatial knowledge based on the entered geo-descriptions.

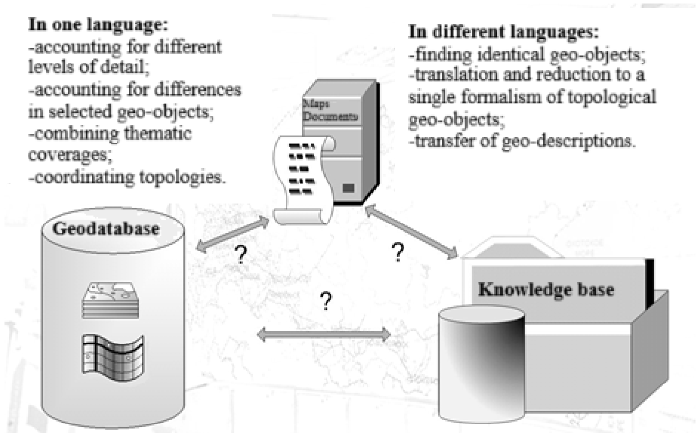
A. A. Lyuty considered three forms of presenting geospatial knowledge: «non-verbal knowledge, verbal knowledge, and knowledge that can be presented in both verbal and non-verbal forms» [3]. Thus, the maps, aerial photographs, lithological and stratigraphic columns, geocharts, paleotectonic schemes, geochemical diagrams, and correlation schemes are nothing more than various geo-text.

One of A. A. Lyuty's most significant contributions to geoinformatics was development of the semiotics of geoinformation, which he called geosemiotics [3]. In geosemiotics, geographic concepts are divided into five categories (Pic. 1).

A. A. Lyuty studied the problem of correlating the figurative and verbal within the framework of the task of geosemantic description of geographical space for development of an electronic image of the Earth. His research on the location and mutual influence of geographical space elements made it possible to solve



Pic. 1. Categories of geographical concepts.



Pic. 2. Tasks of matching geo-descriptions.

geoinformatics problems based on geo-descriptions.

Tasks of Matching Geo-Descriptions

The heterogeneity of geodata causes high costs for development, storage, processing and support of geo-descriptions. Their interoperability causes additional burden due to non-correlated formats and differences in geodata providers and software services. The task of matching geo-descriptions for their joint coordinated use is an important undertaking that requires software, technological and linguistic solutions [4; 5]. For the joint use of geodata based on the matching of geo-descriptions, a semantic level of interoperability is needed, but the autonomy of the developed software systems leads to localisation of the specifics of geo-descriptions. Attempts to integrate heterogeneous geo-descriptions with non-correlated formats lead to costs for coordinating contradictory geodata models that exceed the costs of the development of geo-descriptions itself.

The designed geodatabases are most often of a personalised or departmental nature, which requires additional measures for their joint use. If, at the same time, geodatabases tend to be constantly replenished, then additional requirements arise for scalability of GIS and development of models for matching geo-descriptions that ensure the consistency and integrity of geodata.

Matching geo-descriptions includes procedures for processing references and establishing links between different levels of detail to implement interoperability between the matched geospatial objects [6]. Interoperability is necessary for functional compatibility and correct matching of geodata of different levels of detail.

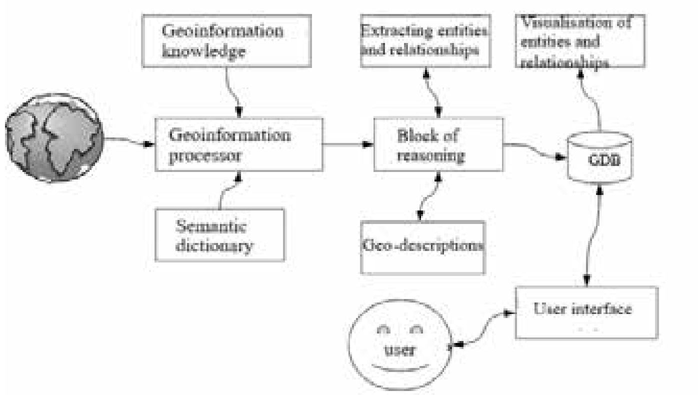
Additional difficulties are caused by the presence of multilingual geo-descriptions, which is not such a rare phenomenon. The aggregate geo-description is built based on the synthesis of geo-descriptions presented in different natural languages. The tasks of matching geo-descriptions in different languages are presented in Pic. 2.

Successful matching of geo-descriptions ensures:

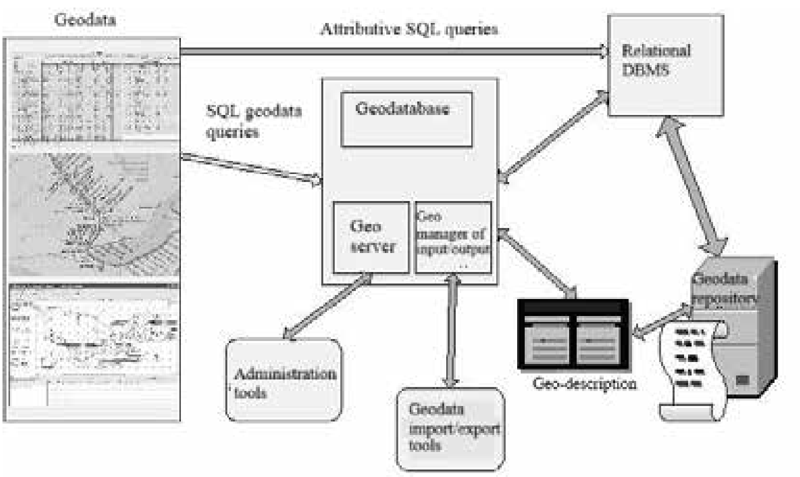
- Improvement of information content.
- Increase in the semantics of geo-descriptions.
- Identification of the quality of geo-descriptions.
- Increased interoperability of expert analysts.
- Testing of methods for matching geospatial objects.
- Development of methods for transforming geo-descriptions.

Formation of Geo-Descriptions

The content of the geoinformation portal is not limited to visual cartographic information. The geodatabase contains and processes a large amount of attributive (semantic) data in the form of different types of texts and documents. The geoinformation portal, as noted by A. A. Lyuty, is an «electronic library with geo-text in different natural languages». The heterogeneity of geodata leads to the semiotic heterogeneity of geo-text stored on the geoinformation portal. From the standpoint of the query-response mode of the database, the processing of existing geo-texts should ensure obtaining the requested information. Pic. 3 shows the blocks that provide



Pic. 3. Analysis and processing of geo-descriptions in the form of geo-texts.



Pic. 4. Geoinformation interaction of geo-descriptions.

analysis and processing of geodata presented by geo-descriptions in the form of geo-texts.

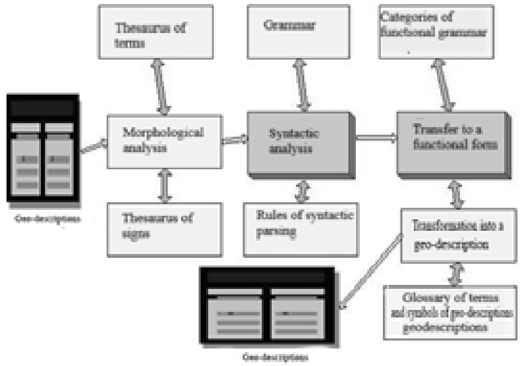
The geoinformation environment for creating and maintaining geo-descriptions is a synergy of the GIS information content, the repository storage environment, and of software aimed at manipulating the necessary geodata. It is worth noting the specifics of the geodata repository, which acts as a coordinator for implementation of the matching of geo-descriptions with other geo-descriptions, ensuring relevant interoperability of geo-descriptions (Pic. 4).

Transformation of Geo-Descriptions

The need to transform geo-descriptions arises if it is necessary to analyse and match multilingual geo-descriptions, synthesising mutually complementary relevant geo-descriptions. It is quite probable that such a transformation is of the greatest importance for formation of a geodata repository, due to a smaller number of constraints on the stored structures compared to a geodatabase.

The linguistic nature of the task of transforming geo-descriptions makes it similar to the task of translating natural language texts using special programs that perform interlingual transformation [7]. If we choose a highly formalised GIS language in which the geo-description is performed, then existing methods of interlingual transformation can be used to transform geo-descriptions.

The key idea of the proposed approach to the analysis of GIS linguistic structures is a simple proposition: to represent grammatical rules as functions in the mathematical sense, that is, as a mapping of the type $A \xrightarrow{f} B$, where A and B are some finite sets. For each of the GIS languages, there is a corresponding data analytics machine, thanks to which grammatical parsing can be performed automatically, i.e. by means of computational procedures. The use of a data analytics machine based on the paradigm of the functional approach assumes conducting the analysis of normalised texts in natural language.



Pic. 5. Blocks of transformation of geo-descriptions.

Normalised text is a text processed by a syntactic and/or morphological analyser. As another constraint, it should be noted that it is necessary to specify the subject area of the analysed geo-text.

The approach to the description of grammar rules as to the assignment of a mathematical function (with its domain of definition and range of value) makes it possible to apply the paradigm of functional programming, which is based on pure functions and immutable data structures, when implementing the data analytics machine in software. It should also be noted that the use of the functional programming environment as a data analytics machine allows the designer to use tuples and «guard functions».

An atom is any uniquely defined identifier. Atoms are intended to specify the attributive characteristics of geolanguage attributes. In this way, all possible attribute characteristics of geolinguistic structures are specified, and the formalism of functional programming turns out to be a sufficient formalism for the analysis of geolinguistic structures, so that there is no need to use any other mathematical tools.

Using a data analytics machine based on a functional programming language as a tool will allow us to design a data analytics machine that implements formal rules for transforming geo-descriptions in the form of mathematical functions.

Based on the grammar of functionally defined geo-texts and choosing a functional programming environment [8; 9], we can obtain a data analytics machine for analysing geo-texts relevant to the selected grammar. Besides, choosing a functional programming environment gives developers access to control at the level of such structures as language atoms and tuples when processing attributive characteristics of geo-descriptions.

Defining attribute properties takes on a different meaning here. When moving to the level of tuples and atoms in a functional language, working with attribute characteristics is no longer the prerogative of the geodatabase; this load is taken on by the functional programming language. Accordingly, support for the integrity and processing of attribute characteristics is carried out using a functional language, and not a geographic information system. This circumstance greatly simplifies the creation of a data analytics machine for transforming geo-descriptions, since attribute characteristics are processed outside the GIS model.

The main advantage of using functionally computable grammars [9] when working with attribute characteristics of geo-descriptions is that the grammar rules are already elements of a functional programming language, which simplifies the solution of a number of issues when designing a geoinformation system, such as resolving ambiguities in the interpretation of rules or setting the rules themselves. At the same time, when setting rules, it is possible to use both the tools of the functional language itself and its extensions, such as «guard functions», which allow you to set rules in a flexible form, which in turn will make working with attribute characteristics of geo-descriptions more adaptable to the problem being solved.

Thus, having formalised attribute characteristics of geo-descriptions and a set of grammar rules specified in a functional form, it is possible to use a functional programming language to process them and not design a separate data analytics machine and also take the processing of rules out of the geoinformation model.

It should be noted that the approach described above allows not only to analyse the attribute characteristics of geo-descriptions, but also to generate them in the form required for further tasks of the system.

The system for transforming geo-descriptions is shown in Pic. 5.

It is easy to see that the processing stages during the transformation of geo-descriptions are similar to the processing stages during interlingual transformation [10; 11]. Morphological and syntactic analysers, as noted above, are necessary to obtain a normalised geo-text, which is transferred to the transfer unit in a functional form. Functional programming tools use grammar rules, synthesising the functional form of a geo-description. The final stage of the resulting geo-description can ensure its translation into another geolanguage.

The proposed scheme for transformation of geo-descriptions is aimed at providing the ability to process geo-descriptions presented in different geolanguages when working with GIS.

CONCLUSION

The article presents an approach that allows considering the semantics of geospatial objects by analysing and transforming geo-descriptions of their attributive characteristics. The possibility of simplifying the processing of attributive characteristics of geodata is based on the use of functional programming languages. The research studies the issues related to the formation, matching and transformation of geo-descriptions for updating geoinformation in the GIS information environment. To implement the transformation of geo-descriptions, a simplification of the formation of a tool is proposed based on the functional programming paradigm for representing geoinformation and a set of geo-texts in various geolanguages.

The implementation of a geo-description transformation system based on a functional programming environment will solve the problem of transforming geo-descriptions similarly to transforming interlingual texts.

As a further study in the field of geo-descriptions, it is suggested to develop a specialised functional programming environment designed to solve the problems of processing and analysing geo-descriptions.

REFERENCES

1. Dulin, S. K., Dulina, N. G., Nikishin, D. A. Problems of maintenance of semantic geo-interoperability and coordination of understanding of geodata semantics. *Systems and Means of Informatics*, 2016, Vol. 26, Iss. 1, pp. 86–108. DOI: 10.14357/08696527160107.

2. Tarasov, V. B. The Problem of Understanding: The Present and Future of Artificial Intelligence [*Problema ponimaniya: nastoyashchee i budushchee iskustvennogo intellekta*]. *Open Semantic Technologies for Intelligent Systems (OSTIS-2015)*: Proceedings of V International scientific-technical conference (Minsk, 19–21 February 2015), 2015, Iss. 5, pp. 25–42. EDN: ZIAXUN.

3. Lyuty, A. A. Map Language: Entity, System and Function [*Yazyk karty: sushchnost, sistema, funktsii*]. Moscow, GEOS publ., 2002, 327 p. ISBN 5-89658-018-5.

4. Shihong Du, Qimin Qin, Qiao Wang, Haijian Ma. Evaluating structural and topological consistency of complex regions with broad boundaries in multi-resolution spatial databases. *Information Sciences*, 2008, Vol. 178, Iss. 1, pp. 52–68. DOI: https://doi.org/10.1016/j.ins.2007.07.023.

5. Schreiber, G. Knowledge acquisition and the web. *International Jpunral of Human-Computer Studies*, 2013, Vol. 71, pp. 206–210. [Electronic resource]: https://www.cs.vu.nl/~guus/papers/Schreiber13a.pdf. Last accessed 17.04.2024.

6. Makarenko, S. I., Oleynikov, A. Y., Chernitskaya, T. E. Models of interoperability assessment for information systems. *Systems of Control, Communication and Security*, 2019, Iss. 4, pp. 215–245. DOI: 10.24411/2410-9916-2019-10408.

7. Kozerenko, E. B. Linguistic modeling for machine translation and knowledge processing systems [*Lingvisticheskoe modelirovanie dlya system mashinnogo perevoda i obrabotki znanii*]. *Informatika i ee primenenie*, 2007, Vol. 1, Iss. 1, pp.54–65. EDN: KVICUX.

8. Dushkin, R. V. Functional Programming in Haskel [*Funktsionalnoe programmirovanie na yazyke Haskel*]. Moscow, DMK Press, 2006, 608 p. ISBN 5-94074-335-8.

9. Dulin, S. K., Ermakov, P. V. Functional-computable grammars in problems of text analysis in natural language [*Funktsionalno-vychislimie grammatiki v zadachakh analiza tekstov na estestvennom yazyke*]. Moscow, CC RAS, 2011, 22 p.

10. Dulina, N., Kozhunova, O. Information monitoring system: a problem of linguistic resources consistency and verification. *Third International Conference «Problems of Cybernetics and Informatics» (PCI'2010). Baku, Elm*, 2010, pp. 56–58.

11. Kozerenko, E. B. Linguistic aspects of informatics [*Lingvisticheskie aspekty informatiki*]. *Systems and means of informatics // Special issue «Scientific and methodological problems of informatics»*, 2006, Iss. 16, pp. 88–111. [Electronic resource]: http://www.ipiran.ru/journal/collected/2006_16a/01–03.pdf. Last accessed 15.02.2024. ●

Information about the authors:

Rozenberg, Igor N., *D.Sc. (Eng), Professor, Corresponding Member of the Russian Academy of Sciences, Chief Scientific Officer of Russian University of Transport, Moscow, Russia, tu@miit.ru.*

Dulin, Sergey K., *D.Sc. (Eng), Professor, Leading Researcher Federal Research Centre «Informatics and Control» of the Russian Academy of Sciences; Chief Researcher of Joint Stock Company «Design & Research Institute for Information Technology, Signalling and Telecommunication on Railway Transport» (JSC «NIIAS»), Moscow, Russia, skdulin@mail.ru.*

Article received 16.02.2024, approved 29.04.2024, accepted 15.05.2024.