Розглядається підхід до вибору геопросторових даних, орієнтований на користувацькі вимоги, а також методи і модель для оцінки якості і вибору наборів даних за допомогою нечіткої логіки. Реалізація запропонованого підходу у вигляді системи дозволила застосувати його для різних ГІС-додатків і продемонструвала позитивні результати врахування потреб користувачів на етапі вибору просторових даних

Ключевые слова: геоінформаційні системи, інформаційне забезпечення, геопросторові дані, оцінка якості, нечітка логіка

Рассматривается подход к выбору геопространственных данных, ориентированный на пользовательские требования, а также методы и модель для оценки качества и выбора наборов данных с помощью нечеткой логики. Реализация предложенного подхода в виде системы позволила применить его для различных ГИС-приложений и продемонстрировала положительные результаты учета потребностей пользователей на этапе выбора данных

Ключевые слова: геоинформационные системы, информационное обеспечение, геопространственные данные, оценка качества, нечеткая логика

UDC 004.9:528
DOI: 10.15587/1729-4061.2016.75514

DEVELOPING A USER-ORIENTED APPROACH TO SELECTION OF GEOSPATIAL DATA BASED ON FUZZY LOGIC

G. Bielcheva
PhD, Assistant
Department of Media Systems and Technologies
Kharkiv National University of radio Electronics
Nauka ave., 14, Kharkiv, Ukraine, 61166
E-mail: annijja@gmail.com

N. Makogon
Assistant*
E-mail: macnat@rambler.ru

N. Manakova
PhD, Assosiate Professor, Head of Department*
E-mail: NatalyM@rambler.ru

*Department of Applied Mathematics and Information Technology
O. M. Beketov National University of urban economy in Kharkiv
Marshala Baganova str., 17, Kharkiv, Ukraine, 61002

1. Introduction

Geo-information systems are a spatial-oriented platform for development of the commercial and open applications of diverse topical directions. And although the geo-information systems (GIS) have been in development for several decades now, nevertheless, interest in this tool is not fading, passing over to the following level of development, which is frequently called the geo-design in contemporary periodicals [1]. The cornerstone of development of both the commercial and open GIS application is creation of information support or information space. By information provision (IP) hereinafter we will understand the sets of geodata in different formats of representation. Information support of GIS-applications is the basis of the work of modules and realization of functional tasks. The complex structure of geodata, which includes a spatial component, is the main distinctive property from other automated systems.

Depending on the purposes and thematics of application, the composition of IP may include thematic maps and plans, photographs, drawings, technical specifications, text descriptions, multimedia components, contact information, geodetic, ecological, statistical data and other various information. But the most significant time expenditures occur at the stage of selection or production of spatial data, which is connected to the need for comparison of information requirements and qualitative characteristics of the data set.

Thus, the selection of applicable sets of spatial data for realization of a particular geo-information project is the most important task, which is carried out in the initial stages of development. In addition to the obvious peculiarities of selection of spatial data, which include availability of the necessary geographical extents, it is important to adequately evaluate the quality of cartographic material, as well as the needs of potential users of future application. Unfortunately, the process of selection of spatial data (SD) with regard to the needs of potential users remains insufficiently studied, which leads to low degree of automation and the significant time costs of this stage.

The mentioned reasons, namely, the importance of information support and essential labor costs for its development, as well as insufficient account of the user criteria of selection, predetermines the relevance of data research.

2. Literature review and problem statement

Attention to the approaches of evaluation of the quality of spatial data was paid to over the entire period of existence of geo-information systems. Thus, the paper [2], dedicated to the penetration of Big Data approaches into the sphere of GIS, states that the standardized methods were developed in geo-informatics for evaluation and description of the data qualitatively and quantitatively. Based on these methods, the frameworks were formed for selection of the applicable data, and then these frameworks were transformed to standard-defining documents.

In Ukraine, valid national standards in the field of geo-informatics are practically lacking. There is a tendency in the introduction of normative documents concerning separate departmental structures; for example, the standard for topographic data [3] was introduced and the requirements for the data in the area of land use are introduced separately [4]. However, the introduction of such documents is not universal, which predetermines the designing of particular standards of quality and leads to chaos in the formats and structure of spatial data.

The need for introduction of normative documentation for standardization of approaches and procedures of the data exchange, evaluation of the quality of spatial data has been long ago understod by the leading developers of GIS and by scientists working in this field. This is specified by the applied nature of development of geo-information projects, which, depending on the different subject matter, requires processing of a large amount of the data in various formats. In particular, International Organization for Standardization ISO developed a series of standards on geographical information, a part of which is dedicated to the quality of spatial data. The ISO ISO/TS 19104 Standard [5] is dedicated to the questions of terminology of geographical information. The ISO 19115 Standard ISO [6] introduces the norms of creation and evaluation of metadata. The ISO 19157 Standard [7] introduced terminological base of quality of spatial data with the elements and subelements of quality being structurally categorized. The forms of cooperation between the user and the data developer are determined, the structure of reports regarding the quality of product for the purpose of introduction to the national and commercial databases.

Two main structural branches of spatial data, as this is presented in the standard [7], are the thematic accuracy and positioning accuracy. In the majority of studies, the methods of control and evaluation are examined separately for positioning or thematic component. Thus, the article [8] proposes universal statistical methods of control of positioning quality. Their universality is in the fact that the proposed approach is applicable to the evaluation of the position quality of data of any measurability, but completely separately from the thematic component.

The statistical methods of evaluation are also frequently used for evaluating the quality of thematic accuracy, for example, the paper [9] uses the Bayesian analysis. In this work, besides the thematic component, the process of evaluation included metadata, but the questions of positioning accuracy are not examined.

Another approach, recommended by the mentioned standards of evaluation of thematic accuracy, is carried out with the help of the matrix of errors (or the matrix of agreement) as, for example, in the paper [10], which presents data in the tabular form and allows estimating correctness of classification of one or another spatial object.

The most productive is the idea of the comprehensive evaluation of the quality of data, considering simultaneously both positioning and thematic accuracy. Thus, the article [11] uses the method of expert estimations and adding the cognitive (thematic) information for evaluating positioning accuracy. The paper [12] uses the method of selection of dimension of the spatial module based on the positioning component of the data for the estimation of thematic accuracy.

In addition to the questions of quality and evaluation of spatial data, the standard [7] presents internal standards for the development, exchange and cataloging of data (from the side of manufacturer); however, external standards, from the side of user, for the selection of spatial data are not standardized. As is emphasized in the article [13], special importance at present is given to the development of the user-oriented methods of selection and evaluation of data, taking into account the process of the large-scale creation of the open sets of spatial data on voluntary basis (Volunteered Geographic Information).

When conducting the analysis of selection SD for IP of GIS-applications, the methods of determining the quantitative information on quality, valid international and national standards in the area of geo-informatics, it is possible to make the following conclusions based on it:

– the process of selection SD with regard to the needs of potential users remains insufficiently studied, which leads to the low degree of automation and the significant time costs of this stage;

– there are in fact no methods of determining the quantitative information on the quality from the side of user, which predetermines the loss of information on the quality and makes the selection of SD difficult.

Furthermore, within the framework of the user-oriented approach, we will consider that the selection of SD by a human is performed based on qualitative categories (but not quantitative assessments), operating with such verbal concepts as “little”, “much”, “fast”, “slow”, “high”, “low” and others while quantitative assessments are only auxiliary in nature. This displacement of focus from the quantitative metrics to the qualitative categories predetermines the selection of fuzzy logic as mathematical apparatus for formalization of the approach. Fuzzy logic makes it possible to describe the parameters of SD quality from the point of view of their semantics and to express the value of the parameters directly in the linguistic scale, which is qualitative in nature and will make it possible to comprehensively process qualitative information and quantitative data.

The main focus of the study is directed toward the study of the methods of calculation of the indicators of SD quality, to the development of the model of the estimation of SD quality, taking into account the user’s requirements for the cartographic material.

3. The purpose and objectives of the study

The purpose of this work is the development of the user-oriented approach of selection of the spatial data, taking into account qualitative requirements of users.

To achieve the set goal, the following tasks were formulated:

– development of the model of estimation of the quality of spatial data, taking into account qualitative categories;

– development of the methods of selection of spatial data, taking into account the user’s requirements on the positioning and thematic accuracy;

– design and program realization of a prototype of the system, which implements the developed user-oriented approach of selection of spatial data and implementation of the designed system in the process of realization of GIS-applications for practical verification of the approach.

4. Methods and model of evaluation of the quality of spatial data taking into account the user's requirements

Let us present the process of formation of information provision of GIS-applications in the form of minimization of the function

$$\Phi_{N_p}^{\text{form.}} = \tau_h\left(\overline{T},\overline{K},R(h),F(h),D(h)\right) \to \min_{\overline{K},R(h),D(h)\in\Omega},$$

time expenses τ_h depend on the totality of factors in the moment h (the moment of formation of information provision), N_p is the sufficient level of fullness of spatial data for realization of GIS-application.

The factors, which influence the process of formation of information provision, include: requirements for the information provision

$$\overline{T} = \left\{T_s,\ T_d,\ T_u\right\},$$

which consider requirements from the side of international and national standards in the area of geo-informatics T_S, from the side of developer – T_D and user – T_U, the indicators of quality of spatial data

$$\overline{K} = \left\{k_1, k_2, \ldots, k_q\right\},\ q = \overline{1,u};$$

internal resources for development of information provision R(h); external factors F(h); economic factors D(h). The zone of constraints Ω includes limitations on the quality of spatial data $\overline{K} \geq \overline{T}$, on internal resources and finances.

Let us form the vector of indicators of quality of spatial data $\overline{K}$, which includes fulness, time accuracy, logical alignment, positioning accuracy and thematic accuracy of spatial data. This set of indicators corresponds to the requirements of international standards on the quality of geographical information and is the basis of the model of evaluation of the quality of spatial data in this study:

$$K \quad f\left(k_1^i,\ k_2^i,\ k_3^i,\ k_4^i,\ k_5^i\right).$$

The requirements to the SD quality T_u^K with regard to existing methods of the GIS-analysis, which are used for realization of FT of GIS-application, are possible to divide into two groups: with the use of the thematic accuracy $T_u^{K(TT)}$ and the positioning accuracy $T_u^{K(PT)}$ of SD (three other indicators of the model are significant for any method of the GIS-analysis). The variative nature of T_u^K predetermines development of the methods of SD selection according to the given value of the indicators.

The qualitative indicators k_4^i and k_5^i are determined with the help of the developed methods, which consider the user's requirements:

$$k_4^i = f_4\left(PT_i,\ VT_i,\ T_u^{k_4}\right),\ k_5^i = f_5\left(l_j, T_u^{k_5}\right)$$

and are assigned to the model's higher level. Then we obtain integral index for the comprehensive assessment of the SD quality K^i in the form:

$$K^i = f_i\left(k_q^i,\ T_u^K\right),$$

where $q = \overline{1,5}$.

4. 1. Method of selection of spatial data according to the indicator of positioning accuracy

Requirements for the positioning accuracy with regard to the user $T_u^{k_4}$ are predetermined by the value of the errors of the planned and high–altitude accuracy ($a_{PT(VT)}$, $b_{PT(VT)}$, $c_{PT(VT)}$). In practice, if the user's requirements coincide with the requirements of standards, the cartographic material needs partial processing ($b_{PT(VT)} = T_s$). The value of the error $a_{PT(VT)}$ is determined experimentally and connected to the use of methods of automatic SD processing in the environment of development of a GIS-application. The value $c_{PT(VT)} \leq T_s$ leads to the unacceptable error, where further SD processing is impossible. The significance of high-altitude accuracy depends on the measurability of cartographic material. In its turn, the value of planned and high-altitude SD accuracy are indicated in the metadata.

If $\overline{x_i} = \left\{SD_i, xml_i\right\}$ is the set of SD sets for IP, where SD_i is the digital SD, xml_i is the XML-file of the metadata, with the description of SD. For provision of the positioning accuracy, it is necessary

$$\left\{PT_i,\ VT_i,\ T_u^{k_4}\right\} \to k_4^i.$$

Let us present the method of SD selection according to the indicator of the positioning accuracy in the form of the sequence of stages:

Stage 1. Formation of the set of SD sets $\overline{x_i}$.

Stage 2. Formation of the vector of the user's requirements $T_u^{k_4}$. Determining the requirements to the measurability of the cartographic material δ (two-dimensional or three-dimensional) of GIS-application and permissible error of the planned and high-altitude accuracy $a_{PT(VT)}$.

Stage 3. Filtration of the SD set $\overline{x_i}$ according to the vector of the user's requirements

$$T_u^{k_4} = \left\{\delta, a_{PT(VT)}\right\}$$

by the set of fuzzy rules in the form

$$\left(PT_i = \tilde{p}_j^1\ \theta_j\ VT_i = \tilde{p}_j^2\ \theta_j\ \delta = \tilde{p}_j^3\right) \Rightarrow k_4^i = \tilde{t}_j^4,\quad j = \overline{1,m},$$

where $\tilde{p}_j$ is the term, for evaluating the variables PT_i, VT_i, δ in the j-th rule; θ_j is the logical operation of the j-th rule. The functions of belonging and the terms of variables are determined. For the planned and high-altitude accuracy $\tilde{p}^{1,2}$ = ={low, medium, high}, for the measurability of cartographic material $\tilde{p}^3$ ={two-dimensional, three-dimensional} and for the output variable $\tilde{t}^4$ ={low, medium, high}. Let us determine the set of fuzzy rules, which describe procedures of processing the input parameters:

$$\left(PT_i = \tilde{p}_1^1 \wedge VT_i = \tilde{p}^2 \wedge \delta = \tilde{p}_1^3\right) \Rightarrow k_4^i = \tilde{t}_1^4,$$

$$\left(PT_i = \tilde{p}_{1,2}^1 \wedge VT_i = \tilde{p}_1^2 \wedge \delta = \tilde{p}_2^3\right) \Rightarrow k_4^i = \tilde{t}_1^4,$$

$$\left(PT_i = \tilde{p}_1^1 \wedge VT_i = \tilde{p}_2^2 \wedge \delta = \tilde{p}_2^3\right) \Rightarrow k_4^i = \tilde{t}_1^4;$$

$$\left(PT_i = \tilde{p}_2^1 \wedge VT_i = \tilde{p}^2 \wedge \delta = \tilde{p}_1^3\right) \Rightarrow k_4^i = \tilde{t}_2^4,$$

$$\left(PT_i = \tilde{p}_1^1 \wedge VT_i = \tilde{p}_3^2 \wedge \delta = \tilde{p}_2^3\right) \Rightarrow k_4^i = \tilde{t}_2^4,$$

$$\left(PT_i = \tilde{p}_2^1 \wedge VT_i = \tilde{p}_{2,3}^2 \wedge \delta = \tilde{p}_2^3\right) \Rightarrow k_4^i = \tilde{t}_2^4,$$

$$\left(PT_i = \tilde{p}_3^1 \wedge VT_i = \tilde{p}_{1,2}^2 \wedge \delta = \tilde{p}_2^3\right) \Rightarrow k_4^i = \tilde{t}_2^4;$$

$$\left(PT_i = \tilde{p}_3^1 \wedge VT_i = \tilde{p}^2 \wedge \delta = \tilde{p}_1^3\right) \Rightarrow k_4^i = \tilde{t}_3^4,$$

$$\left(PT_i = \tilde{p}_3^1 \wedge VT_i = \tilde{p}_3^2 \wedge \delta = \tilde{p}_2^3\right) \Rightarrow k_4^i = \tilde{t}_3^4.$$

Stage 4. Formation of the SD set for IP of GIS-application $\overline{x_i'}$, where $k_4^i = \tilde{t}_3^4$.

The method of SD selection according to the indicator of quality k_4^i implies the possibility of adjustment of tolerance error in the planned and high-altitude accuracy; it considers measurability of the cartographic material and presents the result in the form of expert recommendations, which makes it possible to consider the user's requirements $T_u^{k_4}$.

4. 2. Method of selection of spatial data according to the indicator of thematic accuracy

The requirements to the thematic accuracy $T_u^{k_5}$ with regard to user under conditions of the lack of standard cartographic material are determined by the method of calculation H={L, S, V}, which it is determined by the user depending on the topic of a GIS-application. In practice, with GIS-applications for life support (safety, military, control and guarding of natural resources, control and design of utility networks and communications, etc.), at determining the thematic accuracy, it is necessary to strive for the minimal risk (the S method), connected to the errors in the arrangement of classification features of SD. For the GIS-applications of economic sector (enterprise, business, tourism management, etc.), it is necessary to strive for the maximal gain (the V method). If the topic of a GIS-application is neutral relative to the previous groups, then it is possible to use the weighted medium gain (the L method) during determining a thematic accuracy.

The set of classification features of the cartographic material (l_1, l_2,... l_m) is indicated in the metadata, then the SD thematic accuracy from the side of user is determined as

$$\left\{l_j, T_u^{k_5}\right\} \to k_5^i \quad j = \overline{1,m}.$$

Let us present the method of SD selection by the indicator of thematic accuracy in the form of sequence of the stages.

Stage 1. Formation of the set of SD sets $\overline{x_i}$.

Stage 2. Formation of the user's requirements $T_u^{k_5}$ in the form of the method of determining the indicator H={L, S, V} depending on the topic of a GIS-application.

Stage 3. Filtration of the SD set $\overline{x_i}$, according to the user's requirements $T_u^{k_5}$ with regard to the selected method H by the matrix

$$A = \left(a_{ij}\right), \ i = \overline{1,n}, \ j = \overline{1,m},$$

where $a_{ij} = h_m - l_m$, h_m is the set of probable classification features

$$f(h_m) = \begin{cases} 0, & h_m \le a, \\ \dfrac{h_m - a}{b - a}, & a \le h_m \le b, \\ 1, & h_m \ge b, \end{cases}$$

in the interval [a, b], where a=min l_j, b=max l_j.

According to the selected method H, we determine L_i, S_i, V_i:

$$L_i = \max_{1\le i\le n} \frac{1}{m}\sum_{j=1}^{m} a_{ij}; \quad V_i = \max_{1\le i\le n}\min_{1\le j\le m} a_{ij};$$

$$R = \left(r_{ij}\right), \ i = \overline{1,n}, \ j = \overline{1,m},$$

$$r_{ij} = \max_{1\le i\le n} (a_{ij}) - a_{ij}, \quad S_i = \min_{1\le i\le n}\max_{1\le j\le n} r_{ij}.$$

Stage 4. Formation of the SD set for IP of GIS-applica tion $\overline{x_i'}$, in descending order (ascending) L_i, S_i, V_i, which corresponds to the indicator of the thematic accuracy of cartographic material.

The method of selection of spatial data by the indicator of thematic accuracy is implemented under conditions of the absence of the reference cartographic material and implies selection of the method of calculating the indicator that makes it possible to consider the user's requirements $T_u^{k_5}$.

4. 3. Model of evaluation of the quality of spatial data

Determining the integral indicator of the SD quality includes several stages, which predetermine the hierarchic structure of the model. The procedure of determining the indicators k_4^i and k_5^i is conducted based on the developed methods. The obtained measures of belonging are transferred to the next level of hierarchy, which contains all indicators of quality (time, positioning and thematic accuracy, fullness and logical alignment of spatial data).

The terms and functions of belonging are determined. The dependency of input-output is formed by the fuzzy rules of the form:

$$\left(k_1^i = \tilde{t}_j^1 \, \theta_j \, k_2^i = \tilde{t}_j^2 \, \theta_j \, \ldots \, \theta_j \, k_q^i = \tilde{t}_j^q\right) \Rightarrow K^i = \tilde{d}_j,$$

$$j = \overline{1,m}, \quad q = \overline{1,5}.$$

For the first three indicators of the model we introduced the terms $\tilde{t}^{1,2,3}$ ={low, sufficient}, for the thematic accuracy we use the term – "high", in combination with three quantifiers $\tilde{t}^5$ ={not high, high, very high}, for the indicator of output we introduced three terms and the quantifier $\tilde{D}$= ={is very low, low, medium, high, very high}.

The user's requirements to the SD quality include requirements to all indicators of the model

$$T_u^K = \left\{T_u^{k_{1,2,3}}, T_u^{k_4}, T_u^{k_5}\right\}.$$

For the first three indicators of the user's requirements, the $T_u^{k_{1,2,3}} \ge T_s$, are determined by the international standards of quality and for others they are represented in the form of the sets:

$$T_u^{k_4} = \left\{\delta, a_{PT(VT)}\right\}, \ T_u^{k_5} = \left\{H\right\}.$$

The set of fuzzy rules, which consider the user's requirements with regard to two groups of the GIS–analysis methods ($T_u^{K(TT)}$ and $T_u^{K(PT)}$), is formed in accordance with the expressions:

$$\left(k_1^i = \tilde{t}_1^1 \wedge k_2^i = \tilde{t}^2 \wedge k_3^i = \tilde{t}^3 \wedge k_4^i = \tilde{t}^4 \wedge k_5^i = \tilde{t}^5\right) \Rightarrow K^i = \tilde{d}_1,$$

$$\left(k_1^i = \tilde{t}^1 \wedge k_2^i = \tilde{t}_1^2 \wedge k_3^i = \tilde{t}^3 \wedge k_4^i = \tilde{t}^4 \wedge k_5^i = \tilde{t}^5\right) \Rightarrow K^i = \tilde{d}_1,$$

$$\left(k_1^i = \tilde{t}^1 \wedge k_2^i = \tilde{t}^2 \wedge k_3^i = \tilde{t}_1^3 \wedge k_4^i = \tilde{t}^4 \wedge k_5^i = \tilde{t}^5\right) \Rightarrow K^i = \tilde{d}_1,$$

$$\left(k_1^i = \tilde{t}_2^{\ 1} \wedge k_2^i = \tilde{t}_2^{\ 2} \wedge k_3^i = \tilde{t}_2^3 \wedge k_4^i = \tilde{t}_3^{\ 4} \wedge k_5^i = \tilde{t}_2^{\ 5}\right) \Rightarrow K^i = \tilde{d}_4,$$

$$\left(k_1^i=\tilde{t}_2^{\,1}\wedge k_2^i=\tilde{t}_2^{\,2}\wedge k_3^i=\tilde{t}_2^{3}\wedge k_4^i=\tilde{t}_3^{\,4}\wedge k_5^i=\tilde{t}_3^{\,5}\right)\Rightarrow K^i=\tilde{d}_5.$$

The model's fuzzy rules with regard to the requirements $T_u^{K(TT)}$:

$$\left(k_z^i=\tilde{t}_2^{\,z}\wedge k_4^i=\tilde{t}_1^{\,4}\wedge k_5^i=\tilde{t}^{\,5}\right)\Rightarrow K^i=\tilde{d}_2,$$

$$\left(k_z^i=\tilde{t}_2^{\,z}\wedge k_4^i=\tilde{t}_2^{\,4}\wedge k_5^i=\tilde{t}^{\,5}\right)\Rightarrow K^i=\tilde{d}_3,$$

$$\left(k_z^i=\tilde{t}_2^{\,z}\wedge k_4^i=\tilde{t}_3^{\,4}\wedge k_5^i=\tilde{t}_1^{\,5}\right)\Rightarrow K^i=\tilde{d}_3,\ \ z=1,2,3.$$

The model's fuzzy rules with regard to the requirements $T_u^{K(PT)}$:

$$\left(k_z^i=\tilde{t}_2^{\,z}\wedge k_4^i=\tilde{t}^{\,4}\wedge k_5^i=\tilde{t}_1^{\,5}\right)\Rightarrow K^i=\tilde{d}_2,$$

$$\left(k_z^i=\tilde{t}_2^{\,z}\wedge k_4^i=\tilde{t}_1^{\,4}\wedge k_5^i=\tilde{t}_2^{\,5}\right)\Rightarrow K^i=\tilde{d}_3,$$

$$\left(k_z^i=\tilde{t}_2^{\,z}\wedge k_4^i=\tilde{t}_{2,3}^{\,4}\wedge k_5^i=\tilde{t}_1^{\,5}\right)\Rightarrow K^i=\tilde{d}_3,$$

$$\left(k_z^i=\tilde{t}_2^{\,z}\wedge k_4^i=\tilde{t}_1^{\,4}\wedge k_5^i=\tilde{t}_3^{\,5}\right)\Rightarrow K^i=\tilde{d}_4,$$

$$\left(k_z^i=\tilde{t}_2^{\,z}\wedge k_4^i=\tilde{t}_2^{\,4}\wedge k_5^i=\tilde{t}_{2,3}^{\,5}\right)\Rightarrow K^i=\tilde{d}_4.$$

The integral indicator of quality of spatial data may take five fuzzy values. The value "very low" indicates impossibility of using the data set for IP. All the remaining values allow using the data set for IP depending on T_u^K, however, it is necessary to consider constraints in internal resources and finances of realization of a GIS-application.

The designed model of evaluation of the quality of spatial data, which comprehensively considers the user's requirements, does not depend on the subject area of a GIS-application and fully agrees with a series of international standards of the quality of geographical information. Furthermore, for solving practical problems, the indicators of quality of spatial data are determined at each hierarchical level of the model, which provides for their independent use while solving a wide circle of practical tasks.

5. Results of the study and implementation of the developed methods and models

All the methods proposed above are realized within the framework of user- oriented approach in the course of creating an expert system of evaluation and selection of the sets of geo-spatial data. Let us designate the position of this expert system in the process of development of GIS-applications.

Development of a GIS-application, as a rule, is initiated by the technical task for design, which contains information about the purpose, requirements and functional tasks. The sets of spatial data, existing, commissioned or obtained from the open sources, must be checked for compliance with the list of the required layers of GIS-application in accordance with the functional tasks and requirements to the quality with regard to the user. The implementation of this evaluation of the sets of geodata can be performed with the help of the realized expert system (Fig. 1).

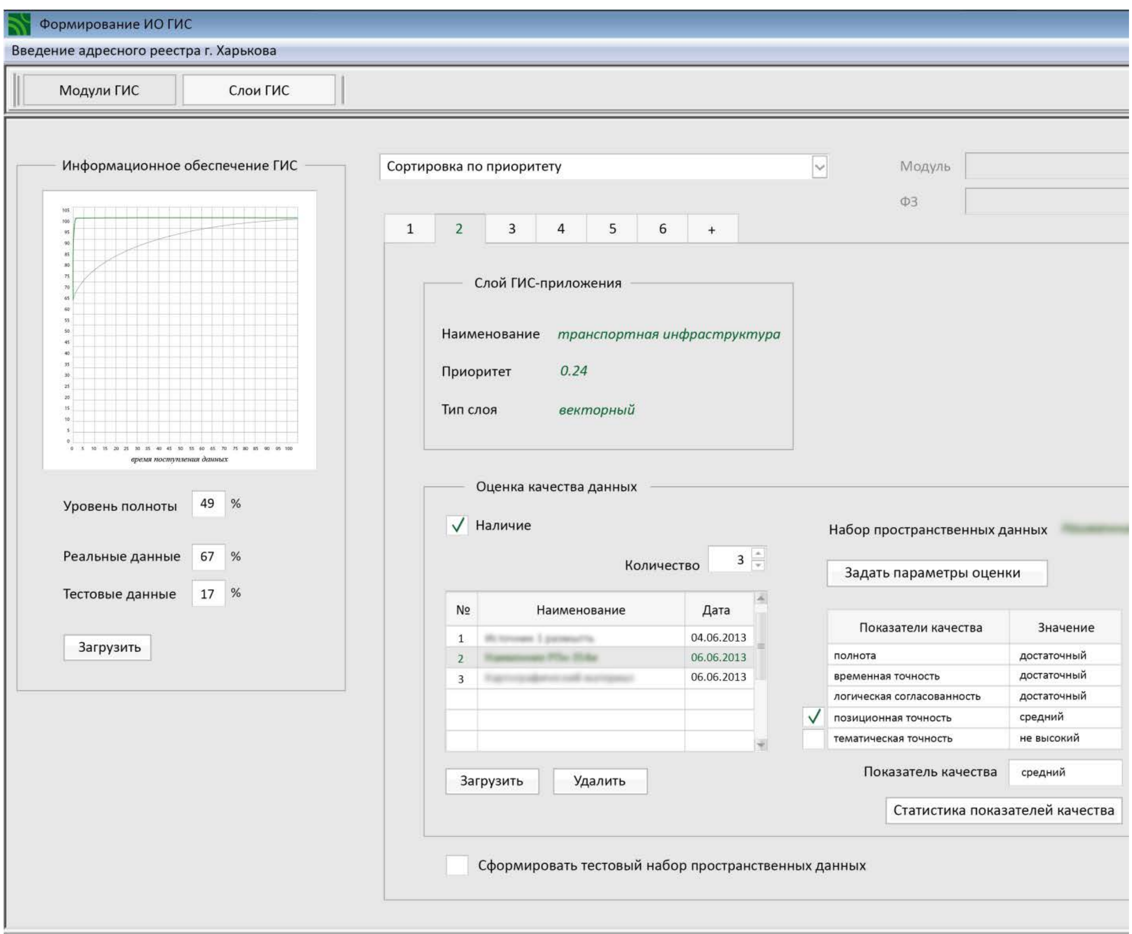

Fig. 1. Forms of registration of the tasks and layers of GIS-application

Fig. 1 presents the form of evaluation of the existing sets of spatial data. By means of the automated form it is possible to indicate compliance of the necessary layers to the user's requirements according to all necessary classification features, prescribed in the technical task. The integrated estimation makes it possible to substantiate decision making on the use of various sets of spatial data.

5. 1. Selection of spatial data in the tasks of monitoring emergency situations

The monitoring of natural and technogenic phenomena is conducted based on operative satellite imaging. During automatic detection of the centers of fire, the data are verified based on the difference of temperatures, which are visually expressed by a change in the values of brightness of pixel in the spectral channels of imaging. To trace the dynamics of fires, to estimate losses and combustion period is possible by using thematic SD of the burntout areas.

Let us determine the possibility of using SD from different manufacturers for monitoring the fire activity. Let us first form the SD sets $\overline{x_i}$. Within the framework of the given research, we used open data on thermal anomalies by the MODIS camera. The photographs of Terra MODIS and Aqua MODIS are taken twice per 24 hours and are the base for detection of the centers of fire. The products of the next level of processing: the fire mask (Fig. 2) and the maximum radiant intensity of fires. For studying the thematic accuracy of spatial data we will apply the first group of the SD sources.

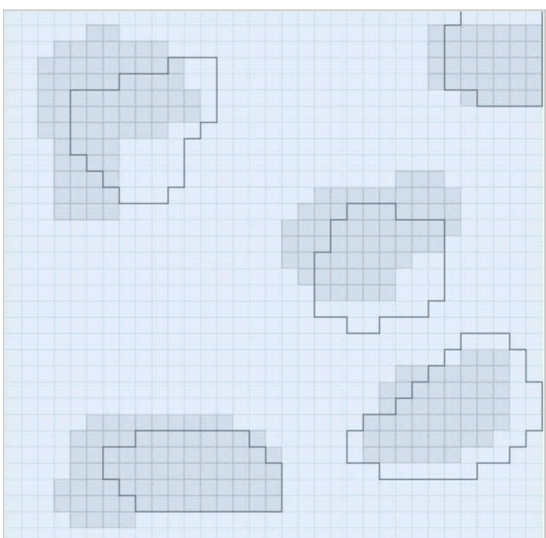

Fig. 2. Fire mask of burnout areas

The subject matter of the solved problem is connected with the GIS of life support, therefore, the method of calculating the indicator of thematic accuracy of SD must ensure the minimum risk, connected with the errors of detection of the centers of fire (the S method). Based on the user's requirements to SD, the calculation of the indicator of thematic accuracy is performed, as well as the selection of cartographic material for the thematic layer of the centers of fire of the GIS monitoring of fire activity (the results of calculation are presented in Table 1).

According to the results of the study, the SD set that is the most approapriate to the user's requirements within the framework of the solved problem – x_4 and x_7. The paper [14] contains a more detailed calculation of the indicator of thematic accuracy of spatial data.

Table 1

Thematic accuracy of SD

No.	SD set	l1	l2	l3	l4	l5	l6	l7	l8	l9	Thematic accuracy (S)	Thematic accuracy level
1	x_1	–26	–18	–66	–32	–1	–11	–65	–31	–28	65	low
2	x_2	–36	–44	–4	–30	–61	–51	–3	–31	–34	60	low
3	x_3	–43	–51	–3	–37	–68	–58	–4	–38	–41	67	low
4	x_4	–8	–16	–32	–2	–33	–23	–31	–3	–6	32	high
5	x_5	–41	–49	–1	–35	–66	–56	–2	–36	–39	65	low
6	x_6	–33	–41	–7	–27	–58	–48	–6	–28	–31	57	medium
7	x_7	–25	–33	–15	–19	–50	–40	–14	–20	–23	49	high

5. 2. Selection of spatial data in the tasks of updating digital maps

Let us determine the possibility of using the photographs of different manufacturers for updating cartographic material of the extensive and sparsely populated terrain (the layers of SD of the forest cover), the scale 1:50000. Let us first form the SD sets $\overline{x_i}$. In practice, in the tasks of updating the maps of extensive and sparsely populated area, it is possible to use panchromatic photographs. Such satellite images can be obtained by the following devices: GeoEye-1, WorldView-1, QuickBird-2, Eros B, Cartosat-2, IKONOS, Kompsat-2, Eros A, Formosat-2, Spot 5, ALOS, Resourcesat-1, IRS 1C/1D, SPOT 2, SPOT 4, Landsat 7 (further in the text satellites' named will be changed the the set of input variables). Two-dimensional cartographic material is used for actualization of the layer of the forest cover. According to the requirements of the standard, the cartographic material of such scale must provide for minimum planned accuracy of 10 m and high-altitude accuracy of 20 m. In practice, within the framework of this research, the requirements for decoding the images using automatic computer processing are slightly higher: the planned – 4 m, and high-altitude accuracy – 17 m.

Fig. 3 demonstrates an example of the results of decoding panchromatic images with the minimally permissable values of the planned accuracy according to the standard, where the color of sectors corresponds to the percentage of the correctly identified spatial objects (the darker the section, the more deviations from the positioning of control points).

Based on the enumerated user's requirements to SD, we calculated the indicator of positioning accuracy and the selection of cartographic material for the layer of the forest cover (Table 2).

Table 2, in addition to the values of positioning accuracy, shows other characteristics of the images, which influence the selection of SD. Within the framework of the solved problem, the desired level of decoding must provide for automatic processing of the cartographic material.

The medium level of positioning accuracy requires additional processing of the photographs in the form of increased frequency of discreteness and interpolation, coordination of histograms, etc., the images with low level of accuracy are not allowed for this task. Due to actualization of the layers with vast territories (604 000 km^2), in addition to high level of decoding, it is necessary to consider the width of the imaging capture and the minimum area of order, then the SD sets x_{13} or x_{15} most fully satisfy the user's requirements within the framework of set task. The paper [15] presents the calculation of the indicator of positioning accuracy of spatial data in more detail.

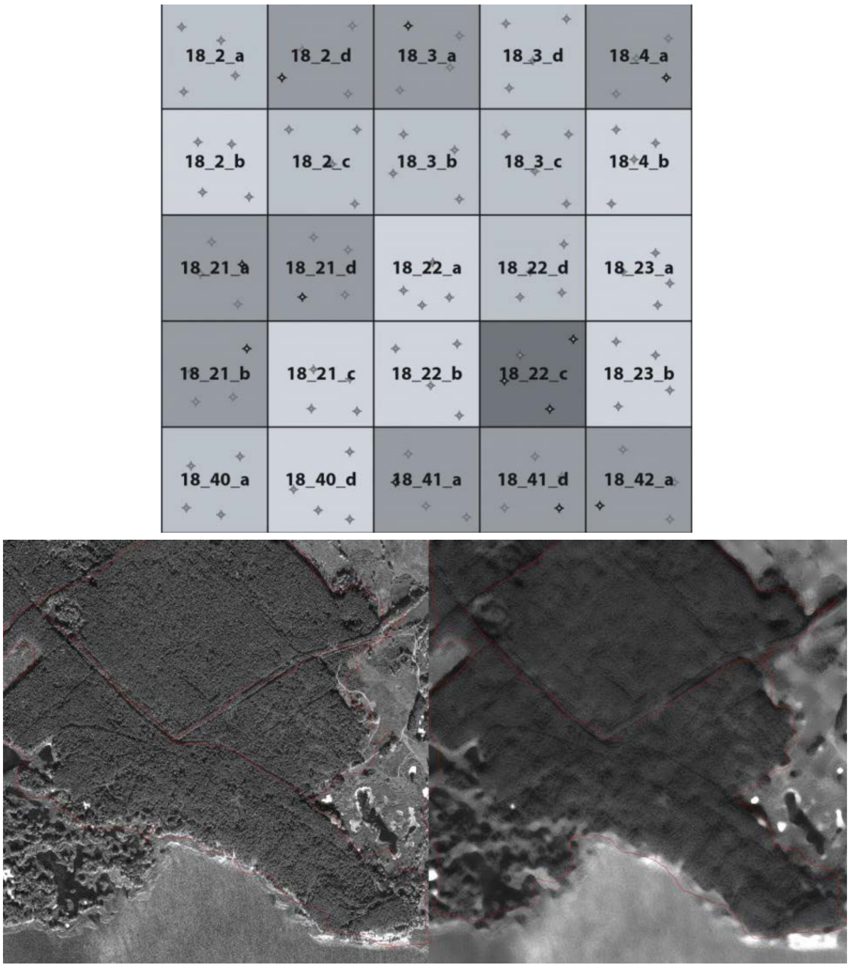

Fig. 3. Errors in the planned accuracy

Table 2

Positioning accuracy of satellite images

No.	1	2	3	4	5	6	7	8	9	10	11	...	17	18
SD set	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	...	x_{17}	x18
Positioning accuracy level	low	medium	low	medium	low	low	medium	high	high	medium	medium	...	high	me dium
Capture width of imaging, km	15,2	17	16,7	7	9,6	11	15	14	24	27	60	...	60	185
Minimal area of order, km^2	49	25	25	15	15	49	50	25	576	183	450	...	900	350000

6. Discussion of the developed approach to the formation of IP and further directions of the studies

The conducted research allowed the authors to form the user-oriented approach and to implement it in the practice of development of GIS-projects. The designed system of evaluation and selection of the sets of geo-spatial data makes it possible to form the information provision of GIS projects. Its verification was carried out at the enterprise "City Information Center" (Kharkov, Ukraine, 2013).

The practical application of the developed approach in the tasks of updating the layers of forest cover and monitoring of fire activity confirmed its reliability and technological efficiency in the process of development of GIS applications. In particular, the selection of spatial data was conducted based on the designed methods, with the use of indicators of quality of positioning and thematic accuracy. The convenience in the formation of the user's requirements in the usual verbal form, due to the application of the methods of fuzzy logic, is also the benfit of the developed approach.

By accumulating practical experience, some shortcomings in the developed approach were revealed. In particular, in the course of verification, it was found that in order to form correct requirements to the quality of positioning accuracy, the users required a consultation from an expert, which slows down its application. This determines certain future direction of the research, namely, while continuing the studies in the methodological part of the approach, it is planned to examine questions of formation of the automated indications of permissible errors in the planned and high-altitude accuracy. Furthermore, we plan to examine possibilities of minimization of time costs for realization of GIS-applications. The period of expectation of receiving the geodata leads to delays in the realization of functional tasks and, in its turn, to postponing the implementation of a project. It is possible to start the stage of realization of functional tasks by introduction of the parallel process of formation of the test spatial data, which is the focus of further direction of research.

7. Conclusions

1. The model of evaluation of the quality of spatial data was developed by using the integral quality indicator, which covers time, positioning and thematic accuracy, fullness and logical alignment of spatial data. The model makes it possible to comprehensively consider the requirements of user during selection of spatial data for the formation of information provision of GIS-applications.

2. The methods of selection of the spatial data according to the indicators of thematic and positioning accuracy were designed. The method of thematic accuracy is implemented under conditions of the lack of the reference cartographic material and includes selection of the method of calculating the indicator, which makes it possible to consider the thematics of a GIS-application. The method of selection of spatial data according to the indicator of the positioning accuracy provides for the possibility of correcting an error of the planned and high-altitude accuracy, considers measuarability of cartographic material and presents the results in the form of expert recommendations, which makes it possible to better consider the user's requirements.

3. The expert system is realized within the framework of the developed methods and the model, combined into a general user-oriented approach. This system makes it possible to form information provision of a GIS-application, and its verification in several projects of different thematics confirmed reliability and technological efficiency in the process of development of actual GIS-projects.

References

1. Wilson, M. W. On the criticality of mapping practices: Geodesign as critical GIS? [Text] / M. W. Wilson // Landscape and Urban Planning. – 2015. – Vol. 142. – P. 226–234. doi: 10.1016/j.landurbplan.2013.12.017
2. Li, S. Geospatial big data handling theory and methods: A review and research challenges [Text] / S. Li, S. Dragicevic, F. A. Castro, M. Sester, S. Winter, A. Coltekin et. al. // Journal of Photogrammetry and Remote Sensing. – 2016. – Vol. 115. – P. 119–133. doi: 10.1016/j.isprsjprs.2015.10.012
3. SOU 742–33739540 0010:2010. Kompleks standartiv. Baza topografichnih danih. Zagal'ni vimogi [Text]. – Kyiv: Minprirodi Ukrainy, 2010. – 27 p.
4. Vymogy do struktury, zmistu ta formatu oformlennja rezul'tativ robit iz zemleustroju v elektronnomu vygljadi (obminnogo fajlu): nakaz Derzh. komitetu Ukrai'ny iz zemel'nyh resursiv № 573 vid 02.11.2009 [Text] // Oficijnyj visnyk Ukrai'ny. – 2010. – Issue 12.
5. ISO/TS 19104:2008. Geographic information – Terminology [Text]. – Switzerland: ISO, 2008. – 102 p. – Available at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=45020
6. ISO 19115:2014. Geographic information – Metadata. Part 1: Fundamentals [Text]. – Switzerland: ISO, 2014. – 167 p. – Available at: https://www.iso.org/obp/ui/#iso:std:iso:19115:-1:ed-1:v1:en
7. ISO 19157:2013. Geographic information – Data quality [Text]. – Switzerland: ISO, 2013. – 146 p. – Available at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32575
8. Ariza-López, F. J. Using International Standards to Control the Positional Quality of Spatial Data [Text] / F. J. Ariza-López, J. Rodríguez-Avi // Photogrammetric Engineering & Remote Sensing. – 2015. – Vol. 81, Isseu 8. – P. 657–668. doi: 10.14358/pers.81.8.657
9. Denham, R. Bayesian analysis of thematic map accuracy data [Text] / R. Denham, K. Mengersen, C. Witte // Remote Sensing of Environment. – 2009. – Vol. 113, Issue 2. – P. 371–379. doi: 10.1016/j.rse.2008.10.008
10. Docan, D. C. Spatial Data Quality Assessment in GIS [Text]: conference / D. C. Docan. – Antalya, 2013. – P. 105–112.
11. Farnood Ahmadi, F. Using cognitive information in the expert interface system for intelligent structuring and quality control of spatial data measured from photogrammetric or remotely sensed images [Text] / F. Farnood Ahmadi, H. Ebadi // Measurement. – 2014. – Vol. 48. – P. 167–172. doi: 10.1016/j.measurement.2013.10.019
12. Stehman, S. V. Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment [Text] / S. V. Stehman, J. D. Wickham // Remote Sensing of Environment. – 2011. – Vol. 115, Issue 12. – P. 3044–3055. doi: 10.1016/j.rse.2011.06.007
13. Kalantari, M. Geospatial Metadata 2.0 – An approach for Volunteered Geographic Information [Text] / M. Kalantari, A. Rajabifard, H. Olfat, I. Williamson // Computers, Environment and Urban Systems. – 2014. – Vol. 48. – P. 35–48. doi: 10.1016/j.compenvurbsys.2014.06.005
14. Bel'cheva, A. V. Teorija igr i raschet pokazatelej jeffektivnosti dannyh [Text] / A. V. Bel'cheva, V. P. Manakov, N. O. Manakova // Radiojelektronika i informatika. – 2011. – Issue 1 (52). – P. 87–89.
15. Bel'cheva, A. V. Fuzzy sets and calculation accuracy of digital maps [Text] / A. V. Bel'cheva, N. O. Manakova // Eastern-European Journal Of Enterprise Technologies. – 2011. – Vol. 3, Issue 2 (51). – P. 29–32. – Available at: http://journals.uran.ua/eejet/article/view/1492/71022