

PAPER • OPEN ACCESS

## Intelligent data model concepts

To cite this article: G M Rudakova *et al* 2019 *J. Phys.: Conf. Ser.* **1399** 033080

View the [article online](#) for updates and enhancements.

A promotional banner for the 240th ECS Meeting. The banner features a colorful striped border at the top. On the left, the ECS logo is displayed in a green circle. To its right, the text reads "240th ECS Meeting" in large blue font, followed by "Oct 10-14, 2021, Orlando, Florida" in a smaller blue font. Below this, it says "Register early and save up to 20% on registration costs" in bold black font, and "Early registration deadline Sep 13" in a smaller black font. At the bottom left, there is a red "REGISTER NOW" button. On the right side of the banner, there is a photograph of a diverse group of people in a professional setting, with a man in a white shirt and tie clapping and smiling in the foreground.

**ECS** **240th ECS Meeting**  
Oct 10-14, 2021, Orlando, Florida  
**Register early and save  
up to 20% on registration costs**  
Early registration deadline Sep 13  
**REGISTER NOW**

# Intelligent data model concepts

G M Rudakova<sup>1</sup>, S L Gladkov<sup>2</sup> and O V Korchevskaya<sup>1</sup>

<sup>1</sup> Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation

<sup>2</sup> LLC Aigeo, 105, office 4-34, 60 years of October ul., Krasnoyarsk, 660079, Russian Federation

E-mail: okfait@gmail.com

**Abstract.** Traditional database systems have a common flaw: the many concepts that make up their metadata must be predefined. By concept we mean concepts, terms, entities. To create a new concept, you must resort to programming or change metadata. There are a number of tasks for the solution of which it is necessary to manipulate metadata. The data mining model should be based on a dynamic concept model. The concept of a data model and a data set, as an interpretation of predicates, is introduced. Based on many basic concepts, many generated concepts are given. The classification, verification, falsification of concepts and the expanded context of many concepts are also considered. Using this approach will provide the ability to control the creation and expansion of distributed information systems.

## 1. Introduction

The address of the property must be subject to strict accounting, otherwise there may be serious difficulties associated with the impossibility of determining the place of residence of citizens or finding objects of enterprises and organizations. Address structure is an information structure that allows you to unambiguously describe the location of the property (to form the address) and to identify the object at a given address of the object.

A serious problem in this area is that the address can be canceled or have an arbitrary value. The address is considered here as a set of values of the requisites. In this regard, it is necessary to formalize the concept of address. At the next stage, it should be normalized. Procedures for the normalization of the values of the requisites and addresses are given in this article. The paper uses the set-theoretic and semantic approaches [1-3].

## 2. Set of basic concepts

The subject area of any database, that is, the set of basic concepts (notions, terms, entities) is a dynamically unchangeable part of the values of its metadata.

The subject area is distinguished from the universal set of concepts by a set of predicates or propositional functions. In Peter Chen's conceptual model "Entity-relationship" [4], concepts are represented as "entities", predicates are "Connections". At the design stage of a logical model, concepts and entities are represented by relationships (tables, nodes, etc.).

The logical scheme erases the differences between the relationships created on the basis of the concepts and the relationships corresponding to the predicates of the domain. But this does not mean



that such differences do not exist, examples of the latter type of relations are “generalization” and “aggregation”, described in detail in the article by John and Diane Smith [5].

Therefore, the set of basic concepts is further considered as a pair, hereinafter referred to as an ontology [6]:

$$O = \langle X, R \rangle;$$

where  $X$  is a finite set of concepts of the subject area, which is represented by the ontology  $O$ ;  $R$  – a finite set of relationships between the concepts of a given subject area.

In addition, we will assume that for the ontology  $O$ , one or several generally valid propositional functions  $C$  are defined, which will be called the context of the concept, group of concepts, and the entire ontology in the following presentation.

### 3. Set of generated concepts

We define the lattice on the set of basic concepts  $X$  as a universal algebra, since this is described in the Birkhoff monograph Theory of Lattices [7].

To do this, we introduce two operations on the set  $X$ : “+” and “-”, which have the following properties:

$$\forall x_i, x_j \mid x_i \in X \text{ and } x_j \in X \implies \exists x_{g+}, x_{g-} \text{ such that, } x_{g+} = x_i + x_j; x_{g-} = x_i - x_j.$$

$$\forall x \in X; x + x = x, x - x = x \text{ (idempotence).}$$

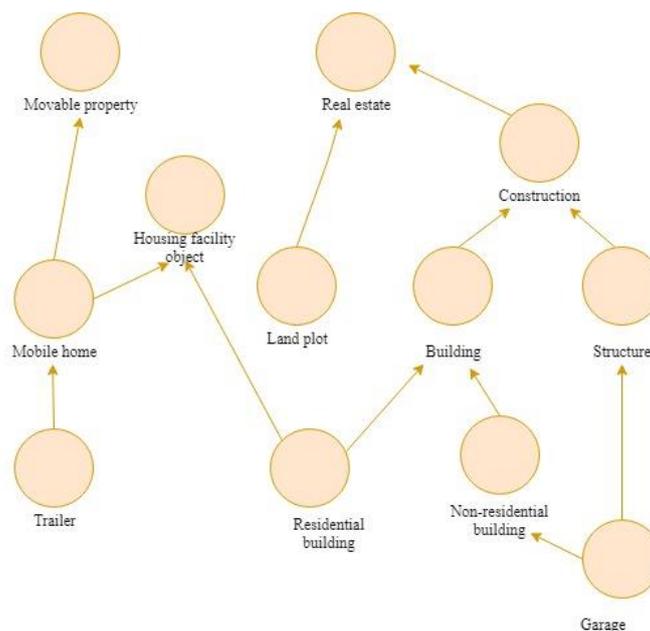
$$\forall x, y \in X; x + y = y + x, x - y = y - x \text{ (commutativity).}$$

$$\forall x, y, z \in X; x + (y + z) = (x + y) + z, x - (y - z) = (x - y) - z \text{ (associativity).}$$

$$\forall x, y \in X; x - (x + y) = x + (x - y) = x \text{ (absorption).}$$

The operation “+” on the set  $X$  is the operation of “generalization” here has the same meaning as in the famous work of John and Diane Smith [5], and the created concept  $x_{g+}$  is called the “generalization”  $x_i, x_j$ .

The purpose of introducing many concepts is to allow the use of the “+” and “-” operations in expressions of the data manipulation language (DML), for example, SELECT \* FROM “Building” + “Construction” (figure1).



**Figure 1.** An example of a deterministic lattice as part intellectual property management model.

The result of this step must match the ontology [6]:

$$O = \langle \bar{X}, \{\vee, \wedge\}, R \rangle,$$

where  $\bar{X}$  is the closure of a finite set of subject area concepts in relation to the operations “ $\vee$ ” and “ $\wedge$ ”;  $R$  – a finite set of relationships between the concepts of a given subject area.

#### 4. Axes and data space

Axis (s) is a generalization of an attribute, as well as a domain in the theory of relational databases. The axis is determined independently of the concepts from the set  $X$ . Each s axis defines a set of  $D_s$  values. The set of axes makes up the universal space  $S = \{s_n \mid 0 < n \leq N\}$ .

Remark 1. The set of axes  $S$  is a subset of  $X$ , i.e., each axis is a concept to which, in particular, the operations “+” and “-” can be applied. The goal of defining the axis as a concept will become especially clear after considering the sequence of concept generation (Conceptual semantics).

The data are points of the space  $S$ . Each point is a set of values corresponding to the axes of the space:  $\{d_{ij} \mid \forall j d_{ij} \in D_i\}$ .

We define on the set  $\bar{X}$  the relation, which each concept is assigned to the subspace  $S_x \subseteq S$ .  $S_x$  is the space on which the  $x$  data of the concept is defined. Pair  $\{x; S_x\}$  is equivalent to a relational algebra relation, where  $x$  is the name of the relation, and  $\{x s_{xi}\}$  is the set of attributes.

To designate a relationship from the set of concepts  $\bar{X}$  to the set of data  $D$  of the universal space  $S$ , we will use the sign of the operation “composition” to combine the concept and the corresponding subspace of the axes  $x S_x = D_x$ .

Let the compositions  $x_1 S_{x_1}$  and  $x_2 S_{x_2}$  be defined, then the subspace that corresponds to the generalization  $x_1 + x_2$  can be obtained by intersecting the subspaces of the original concepts, i.e.

$$S_{x_1+x_2} = S_{x_1} \cap S_{x_2}.$$

Let us explain how the “composition” is performed with  $x_1 + x_2 S_{x_1+x_2}$ . To do this, each axis is associated with a “property” or “method” of the mapping, which corresponds to the relational functions of the grouping  $SUM()$ ,  $MAX()$ ,  $MIN()$ ,  $AVG()$ . So, if the “quantity” axis corresponds to the “property”  $SUM()$ , then the “composition” “computers + tables” \* “number” is converted into the relation “computers” + “tables” \* “number”. Each instance (tuple) of a dynamically created relation will be the sum of the values “computers” \* “quantity” + “tables” \* “quantity”.

The result of this step should correspond to the ontology:

$$O = \langle \bar{X}, \{\vee, \wedge\}, R \rangle,$$

where  $\bar{X}$  is the closure of a finite set of domain concepts in relation to the operations “ $\vee$ ” and “ $\wedge$ ”;  $R$  – a finite set of relationships between the concepts of a given subject area.

#### 5. Classification, verification, falsification of concepts

Introducing the space  $S$ , we set the mapping of the set  $\bar{X}$  to the set of values  $D_S \forall x \in \bar{X} \exists S \Rightarrow S_x(x) = D_x$ .

Now consider the inverse problem: finding the concept  $x \in X$ , which corresponds to the point  $d \in D$ .

Tasks of this type often arise in the Data Mining process, for example, finding stopping places (stops) of a vehicle as a result of analyzing a track from a GLONAS (GPS) receiver, in the process of information exchange.

To solve this problem, we introduce the concept of the membership function

$$\mu_x(d) \in [0 \dots 1] \mid d \in D \vee x \in X,$$

which points to the data set points correspond to the probability that a given point  $d$  belongs to concept  $x$ .

Since each concept  $x$  is associated with a subspace  $S_x$ , then all points outside this space will a priori not correspond to the concept, that is,  $\mu_x(d) = 0 \mid \forall d \in D_x$ .

Let  $S_x$  be such a minimal subspace  $I_x$  such that the membership function takes the same value at the point  $d_i$  of the subspace  $I_x$ , and at each point  $S_x$  whose projection on  $I_x$  is equal to  $d_i$ .

Thus,  $I_x$  is the minimum space of the discriminating data for concept  $x$ .

Then the classification of a point  $d \in D$  will be understood as the problem of correlating it with one or several concepts  $x$  using the set of membership functions  $\{\mu_x(d) \mid d \in I\}$ .

Verification of the point  $d$  of the subspace  $S_x$  will be understood as the task of checking that this point corresponds to the concept  $x$  using the membership function.

The falsification of the point  $d$  of the subspace  $S_x$  will be understood as the task of finding and making such minimal changes in  $d$  that the field of such a transformation does not correspond to the concept  $x$ .

## 6. Conclusion

In contrast to traditional data models, an intelligent data model should be based on a dynamic concept model that allows you to transform existing and create new concepts using “ $\vee$ ” and “ $\wedge$ ” operations. This property will ensure the adaptability of the model both to changes in the meaning of concepts with time and to fundamental changes in external objects, data about which are stored in the information system.

The concepts of an intellectual data model should have verifiability and falsifiability properties. In other words, an intelligent data model should contain a mechanism for checking the correctness of the correlation of external data with a particular concept throughout its use. This property is especially important for an information system intended for storing and processing scientific and historical data.

The concepts of an intellectual data model should have verifiability and falsifiability properties. In other words, an intelligent data model should contain a mechanism for checking the correctness of the correlation of external data with a particular concept throughout its use. This property is especially important for an information system intended for storing and processing scientific and historical data.

The concept definition of an intelligent data model should contain the context of the concept, as a mechanism for combining concepts into sets. This property will provide the ability to manage the creation and expansion of distributed information systems with an intelligent data model, as well as organizing the interaction of information systems with an intelligent data model.

## References

- [1] Gottlob F 2000 *Logic and Logical Semantics* (Moscow: Aspect Press)
- [2] Emmon B 2010 *Informal Lectures on Formal Semantics* (Moscow: Book House «LIBROKOM»)
- [3] Beer S 2005 *Brain Of The Firm* (Moscow: Publisher: Editorial URSS)
- [4] Chen P P 1976 The entity-relational model. Toward a unified view of data *ACM TODS* **1**
- [5] Smith J M, Smith D C 1977 Database Abstractions: Aggregation and Generalization *ACM Transactions on Database Systems* **2** (2)
- [6] Bolotova L S 2012 *Artificial intelligence systems: knowledge-based models and technologies* (Moscow: Finance and Statistics)
- [7] Birkhoff G 1984 *Theory of lattices* (Moscow: Science)