

The method of abstraction in the design of databases and the interoperability

Nikolay Yakovlev^{1*}

¹Moscow State University of Civil Engineering, Yaroslavskoe shosse, 26, Moscow, 129337, Russia

Abstract. When designing the database structure oriented to the contents of indicators presented in the documents and communications subject area. First, the method of abstraction is applied by expansion of the indices of new, artificially constructed abstract concepts. The use of abstract concepts allows to avoid registration of relations many-to-many. For this reason, when built using abstract concepts, demonstrate greater stability in the processes. The example abstract concepts to address structure – a unique house number. Second, the method of abstraction can be used in the transformation of concepts by omitting some attributes that are unnecessary for solving certain classes of problems. Data processing associated with the amended concepts is more simple without losing the possibility of solving the considered classes of problems. For example, the concept "street" loses the binding to the land. The content of the modified concept of "street" are only the relations of the houses to the declared name. For most accounting tasks and ensure communication is enough.

1 Introduction

Still J. Forrester [1] defined the city as urbanized territory, i.e. a system in which interact with different types of business, residential and non-residential people. A detailed description of these components and check the related facts and events are executed in the units of the authorities and city management. An integral part of the functioning of these units is now becoming the operation of automated information systems (AIS) or the automated control systems (management information system). Information support of automated systems, database structures are created on the basis of a set of indicators appearing in the documents. At the beginning of creation of the automated systems in the city of Moscow was made and in some cases specifically stipulated by the customer, when you build to describe the logical structure of the respective databases (at the level of global logical description [2]) most accurately reflect ideas about the subject area, formed in the non-automated control system and reflected in the relevant documents. It can be shown that this approach often leads to difficulties of identification and subsequent integration. These difficulties are due to the natural active role of people in processes of functioning of the city and the nature of these, reflecting the complex activity.

* Corresponding author: yakovlev50@inbox.ru

The data about real objects are often characterized by a plurality of representations. For example, the address of a building in the city of Moscow can consist of the following parts: street; Number and index of the house; Number and index of the hull; Number of the building. Historically, the city was a collection of land plots - estates with buildings on them. The numbering of buildings on the land plot clockwise still takes place in the internal documentation of the Moscow City Bureau of Technical Inventory. From there, the numbering of buildings sometimes penetrates into real addresses. The concept of "hull" is more associated with the processes of mass construction, is not tied to the land and reflects the sequence of construction of nearby buildings. Buildings later built have a larger hull number. In some areas of the city, for example in Zelenograd, often there is no house number, it is replaced by the number of the hull, and part of the hull number is the number of the neighborhood. In addition, the structure can have addresses on several streets, often two for corner buildings at intersections. The streets are subject to renaming. Most often this happens when you replace "building" addresses with "mail" addresses. Depending on the age of the document, the street name may correspond to the position before or after renaming.

In these conditions, the addresses of the same structure in different documents could differ quite significantly. And if the information system required that information on the same structure should be combined, and the data on the structure came from different sources, then the solution of such tasks involves a significant amount of identification work.

Therefore, the phenomenon that is required to deal, lies in the multiplicity of representations of the locations virtually any structures in the city. The building has multiple addresses. And if we have some address yet to be identified, then this address will have to compare with multiple sets of addresses each set of addresses belongs to a certain building.

Traditional development of information support involves the use of objects and their attributes from the subject area [3,4,5,6,7,8,9]. While implicitly it is assumed that for the successful operation of the future system this material enough. Nothing beyond existing information to add not required. The identification procedure can be quite complicated [10]

The second phenomenon that needs to deal with is a varied territorial division of the city, allowing not less than a variety of ways of its representation in information security. The division of the city into spheres of influence of the territorial units of the urban structures is very diverse. Describe the parts of the territorial division can be performed in a verbal description of boundaries or a depiction of the boundaries on paper or electronic maps and charts, or, listing the address lists of the serviced buildings. The situation with borders is sometimes compounded by the fact that not all land within the city is divided into local area. View the land through the list of addresses relevant to the plot of the buildings is complicated by the multiplicity of addresses for each building.

The choice of method of representation of concepts depends on the nature of the tasks we are going to solve with the use of this concept. For most accounts, search tasks and identification tasks, the unit of territorial division on the map is not important. Since the activities of legal entities and individuals in the city is concentrated in buildings, and it's relation to particular structures to unit territorial division. A lot of problems to be addressed in an automated way connected with the use within a single city service information prepared in another city service with the features of territorial division within each service. The translation task pairing such information into language representations on the territorial division of the city leads to the processing of data about land plots, their borders, and mutual complementarity. Such processing is characteristic of geographic information systems.

2 Methods

In order to cope with the variety of locations, designing the corresponding concept applied a variant of the method of abstraction based on the following features. On the one hand, the

applied abstraction to the maximum technically possible. On the other hand, after you build the required concepts are applied the most complete Supplement of this concept to those items that were discarded in the initial abstraction. A new concept has been called "Unique house number" (UDOM) and treated any individual structure. When you build this concept applied principles:

- 1) primary code structure needs to be abstract, i.e., do not depend on any physical parameters, location and similar characteristics of relations of the structure on site
- 2) filling of the database must ensure that the maximum technically possible completeness of data on all buildings in the city
- 3) filling of the database must ensure that the maximum technically possible completeness of data on all addresses of each of the buildings

For solving problems related to the diversity of territorial division of the city, in order to avoid the necessity of processing data on land parcels, used the method of abstraction. In this case, abstracting themselves from the land which divided the territory of the city. After this abstraction remains such an attitude: "the name territorial unit" - the list of unique house numbers belonging to a given territorial unit. In terms of the logical data description is most convenient to represent these relations using certain facets assigned to unique house numbers. Each facet corresponds to a certain territorial division of the city. Since the activities of legal persons connected with the placement in the buildings, with addresses, the idea of productively when considering the territorial aspects, especially if analyzing, and processing the exposed crossing of the territorial divisions.

3 Results

3.1. Unique house numbers and addresses

In 1987-1989 on the basis of hardware and software of automated system for accounting and distribution of residential area in Moscow (management information system "KURS") was established on the basis of a future Database, "The Housing and uninhabited Fund" (for the Moscow city Bureau of technical inventory (MosgorBTI)). In the description of the data in the Statistical card for the building was applied the first two postulates.

The implementation of the first postulate was ensured by the automatic generation (the first time you enter the corresponding statistical card) the so-called "unique house number" (UDOM), which is then not changed throughout the life cycle of the buildings. That is, even after the demolition of buildings and construction on this place another house archival data on demolished structure in the database were contacted by former unique house number.

The second postulate was provided by enter in the database information about all included in MosgorBTI buildings in the city of Moscow and the further maintenance of the database. Note that the entry in the database "The Housing and uninhabited Fund" information about all apartments in all buildings in the city and about all areas all the apartments were made in 1990-1993. Thus, in the machine information base, information about a particular structure is associated with the corresponding unique house number. The ability to restore the original form of the document, from where a specific address came, is ensured by storing in a special table the connection between the unique house number and the entire set of real addresses associated with this building.

The transition from the usual address, which occurs in documents, to the unique number of the house is carried out through identification procedures. These procedures are performed either interactively with direct participation of the operator, or in batch mode, when processing of an array of data with addresses received from another information system is required. If it is not possible to uniquely identify interactively, the document where the

address is found with multiple interpretations is returned to the source for clarification. This situation arises if the database has two buildings with one house number and buildings 1 and 2, in the address for identification there is only the house number, the hull number and the building number are missing altogether.

When identifying in batch mode, man-machine procedures are used with sequential refinement of SQL queries and a gradual increase in the percentage of identified addresses. The implementation of the second principle is very useful and allows increasing the share of the identified addresses. It is important to note that the identification for the address data is performed once - the first time they are entered into the system. In further transformations of data, a unique house number participates. Thus, the principle of one-time data entry into the system is realized. The third principle is used when expanding the use of unique numbers to other automated systems, and ensuring interoperability. The result of the application of this principle increases the efficiency of identification processes.

3.2. The interaction of automated system

The subject of interaction between systems is quite extensive. We will discuss only the question of the use in one system of information about a certain object or phenomenon from another system in those cases when the concepts of the same subject or phenomenon in different systems are different. For example, let it be information about apartments in the same house, whose addresses in interacting systems are represented in different ways: in one system this will be the mailing address, and in another it will be the address of the technical inventory office. In this case, the addresses of the same structure in different systems may not coincide. Large companies that deal with interoperability issues, in particular, offer mutual agreement of interfaces, data formats and exchange protocols. In our case, this is not enough, since the content difference in addresses will not disappear anywhere. Staying at the level of universal methods of ensuring interaction before taking into account specific differences and to ways of overcoming them will not reach you.

Inclusion in the interaction process of abstract concepts that unambiguously represent the subject of interaction is one of the ways to actually ensure the interaction of systems that have multiple ambiguous data about the same subject. We give a number of examples of the implementation of this proposal. For the first time, unique house numbers were generated and applied in the database "Housing and Uninhabited Fund", created for the Moscow City Bureau of Technical Inventory.

The easiest from the point of view of the effectiveness of data exchange, in which information about buildings is involved, would be an option when both systems have a binding to unique house numbers. And such approaches have been organized. When launching an automated social security management system for Moscow residents, information was provided on addresses linked to unique house numbers. In the future, this system provided data to the tax authorities with information about pensioners who were entitled to property tax relief. Since the data was provided with unique house numbers, there were no problems with the identification of addresses.

For the tax authorities in the city of Moscow in 1999 - 2004 was created an automated information system (AIS) "Tax 2 Moscow". For the subsystems of this system, related to the property tax, a part of the database containing information about the buildings was based on unique house numbers. When transferring to the tax authorities from MosgorBTI assessments of apartments for tax purposes, which also contain unique house numbers, again there were no problems with identification. The importance of obtaining "good", tied data with the appraisals of apartments is due to their application both at the annual tax assessment and throughout the year with the direct appeal of taxpayers to the inspection.

For the postal service of the city of Moscow (Glavmospochtamp), a binding of their directory with information on the buildings on the Database "Housing and Uninhabited Fund" with the use of unique house numbers was made. Along the way, the guide itself was refined, which was helped by the realized postulate of the completeness of the data. Within the AIS "Tax 2 Moscow" it was required to solve the problem of mass printing and dispatch of notices on property taxes in such a way that by mail the notices came within the allotted time. Transferred to the tax authorities from Glavmospochtampa information about the buildings contained a unique house number and postal code. As a result, there were no problems with identification and the postal service began receiving in mass order from the tax authorities letters with documents on property taxes, issued by all postal rules and with correct indices.

Feature of the city of Moscow is a strong departmental disunity. Each information system is created taking into account the established traditions and categorically does not want to change itself to facilitate interaction with other systems. For those cases when the direct inclusion of unique house numbers in the operating information system did not result in the creation of an external service that received data from the sending system, processing them with identification and binding to unique house numbers and subsequent transfer of replenished data to the user system. This situation arose with data on the deceased, processed in the information system of the registry office. The fact is that it is impossible to send documents on property taxes to deceased people, every mistake in such a case is fraught with a scandal, and there were no unique house numbers at the addresses of registration in the database of the registrar's offices. This external processing of data from the registry offices solved the identification problem.

3.3 Abstract concepts in the tasks of territorial division of the city.

When describing the units of the territorial division of the city, we will abstract from the configuration of the boundaries and topology of parts of the land. Then for the representation of a specific territorial unit there is only a list of unique house numbers pertaining to the territorial unit in question. And this representation is quite enough for solving a wide range of tasks of automated systems. The implementation of this approach is fairly simple due to the high level of abstraction. At the logical level, a flat table is defined, containing a unique house number and signs of belonging to the corresponding units of territorial division. Let us give some examples.

A street is a collection of buildings belonging to a street. In this case, the structure can belong to different streets. This is the case for corner houses, taking into account the renaming of streets, etc.

The district (according to the old division of the city before perestroika) is a set of buildings that form the district.

The administrative district is a set of buildings forming the administrative district.

The municipal district is the totality of the buildings forming the municipal district.

The administrative district is a set of buildings forming the administrative district.

The municipal district is the totality of the buildings forming the municipal district.

The area of responsibility of the police stations is the totality of the buildings to which the responsibility of the relevant branch is covered.

And so on.

"Layout" of buildings with codes, for example, territorial tax inspections is a routine work, which is sometimes greatly facilitated by knowledge of specific features. So the territory of the "marginal" territorial tax inspectorates (Nos. 11 and on) basically coincided with the old district division of the city, which allowed to reduce the time for the primary marking of the buildings.

We will show the application of this approach for the tasks of processing data on taxes on the property of individuals. During the transfer of data to the tax authorities on the availability of real property subject to taxation from physical persons, identification of addresses of buildings in which there are owned apartments is performed, and binding of addresses to unique house numbers. After processing the data on the owners together with the information on pensioners for accounting for tax benefits, then together with information about the deceased and information on the cost of apartments for taxation, it is required to transfer the data to the territorial tax inspectorates. It is not difficult to isolate the data for each inspection, since for each unique number of the house there is its relation to the corresponding territorial tax inspection. Further, having received the documents with assessed taxes in the territorial inspection, they must be printed and packed in envelopes with the indication of postal codes, which is provided by the presence of a unique home number connection with the postal code of the liaison office. This same connection allows you to solve the problem of possibly even loading the post offices delivery of tax notices in order to meet the temporary restrictions on the transfer of documents.

4 Discussion

The application of the above abstraction is not necessary, but ignoring them can lead to undesirable consequences. When created in 1991-1992 the automated system of the present Department of Property of the City of Moscow developers, it was decided to consider in the database ' structures solely as they appear in the input documents. It was motivated by the fact that the relevant legislation requires that documents in the information system be stored in an identical paper document with the volume and composition of information. As a result, for three months of operation of the system in the database, building addresses accumulated, more than twice the number of all buildings in the city. Under these conditions, the task of identifying data to collect information even for one structure, for example, for resettlement before demolition or reconstruction, often led to the need to refer to paper archives. In a few years, the Department entered into a contract to bind the accumulated addresses with unique house numbers and identification was performed.

5 Conclusions

In the case when information about the same object have many views, how many different addresses of the same structure, it is advisable by abstraction to create a new concept. Implementation in the information system of this concept must be associated with the maximum achievable diversity of the cast when abstracting data. Design and use abstract concepts in problems of functioning of automated information systems and problems of interoperability of information systems in the context of multiple representations of the same objects showed its effectiveness.

Not all data is formally associated with complex objects, for example, land must be handled with appropriate difficulty. Application of the method of abstraction allows you to simplify the data representation and processing. Simplicity, in turn, reduces the likelihood of errors. Of course, this is true for a certain group of tasks. The method is tested in the interoperability of automated systems.

References

1. J. Forrester. Dynamics of the city development. - M, Progress, 1974
2. J. Martin. Organization of databases in computer systems. -M, World, 1980

3. C.J. Date Introduction to Database Systems An 8 Williams (2016)
4. Devacademy.ru (<http://devacademy.ru/posts/sqlite-vs-mysql-vs-postgresql>)
5. M. A. Poltavtseva Storing complex data structures in relational databases. PhD Thesis. (Tver 2008)
6. S.V. Mescherjakov, V. M. Ivanov Effective technology of information systems (SPb: Publishing house Politehnika, 2005) <http://www.polytechnics.ni/>
7. S.V. Mescherjakov, V. M. Ivanov Methods of optimal designing of databases of production equipment - SPb: Publishing house Polytechnic University,2012. <http://gpupress.ru/>
8. T. Halpin, *Information Modeling and Relational Databases: from conceptual analysis to logical design* (Morgan-Kaufmann, San Francisco, 2001)
9. Data Modeling and Database Design (Course Technology, 2007)
10. N. S. Nikolaeva, A. A. Shilnikova, M. V. Yurova Automation, mechatronics and information technologies: materials of the I International scientific and technical, Omsk: Omsk State Technical University (2010)