Data Transfer and Objects Recognition in Building Design

Prof. Dr. A. Pavlov, Moscow State Civil Engineering University (aspavlov@mtu-net.ru)

1 Summary

The problem of data interoperability is now very important. The formal description of construction systems and objects must base upon the modeling for the description of construction data domain. The XML-language was selected as a basis of a universal data format, ensuring natural hierarchy of objects, flexibility, good layout and expandability. The language, developed by the author, is called Building Object Description Extensible Markup Language (bodXML).

The types of all objects used by data transfer should be definite beforehand with existing methods of programming. It limits the possibilities of IT in application of new types. But the recipient software must recognize the building objects even if the kind of object is unknown at the outset. The author offers a set of main topological and geometric properties being sufficient for recognition of main three-dimensional building constructions with flat edges. The tests of artificial neuron network have shown that the recognition of a kind of the constructions represented as a set of indicated parameters happens enough confidently.

2 The analysis of methods of information processing in building

2.1 The tendencies of development of the IT in building design

The various information technologies (IT) are widely applied at present in practice of building design. The main kinds of software, applied in building designing, are: the universal graphic editors, architectural and designer systems (CAD), static analysis systems (FEM), geo-information systems (GIS), facility management systems (FM), organizational and economic systems, and also combine systems. The complex solutions ensuring the accurate transfer of information from one program to other on a technological line-up have the greatest market demands. At the same time the use of a number of the software successfully deciding the local concrete problems is pertinent too. The problem of accurate data transfer and data transformations sharply came up in this case.

2.2 The analysis of methods of information transfer

Huge amount of information is produced for each building project; however the large part of useful information is lost or inaccessible to remaining participants. A lot of information is unduly duplicated. On the contrary as a result there is appearance of distinguishing versions of document. The computer programs often are used independently. Thus, data in this case are transmitted manually. The existing methods of data transformation with standard data formats frequently conduct to loss of semantic information during transfer from one workstation to other because the various programs have different constitution of transmitted data.

The special attention should be given to a problem of interoperability of information technologies. The modern technology of the design documentation preparation is impossible without regular data transfer from one workstation to another. Thereby checking, comparison and transformation of data are necessary. It is necessary, on the one hand, to create scientific fundamentals for construction of the interoperability protocols and, on the other hand, to develop methods of interaction of the software.

The attention of the developers and scientists in this field centered round the problem of accurate data transfer within the framework of information technologies. The main tendency has become aspiration for development the universal language. It would reflect mapped the outcome data of IT, and would read easily by the computer. The particularity of IT in construction is the processing of both graphic and alphanumeric data. There are also other unsolved problems, but the main direction in a solution of compatibility consists in development of the universal format, witch is useful for the description of various data types.

3 The organization of information exchange and data conversion

3.1 System approach to the data transfer problem

The major system problems are concentrated on junctions of interacting systems. It is the prime field for application of a systems engineering. The author offers to distinguish the "infra-design" (or designing of the infrastructure) as the stage of organizational design of complicated project and building systems e.g. a stage of IT subsystems correlations design. The following functional components of a designed IT infrastructure are distinguished:

- Perception and primary transformation of project information;
- Intermediate storage of information in an independent format;
- Interactive enrichment of information with creation of the context-sensitive medium of the project;
- Support of databases containing all necessary information;
- Recognition of objects not accompanied with relevant information;
- Check of the co-ordination of project data;
- Data preparation for other subsystems and external systems;
- Support of project management systems, including data authorization and security.

3.2 The documents in building design

The classification of the documents, which are used on various stages of building object life cycle, is given on purpose, content, source, carrier and structure of information. For instance, the sound recordings, text documents, forms, tables, illustrated texts, graphics images, animations, complex structured documents, databases, knowledge bases, and also models are segregated on structure of information. The main kinds of information, for example, in the text documents and tables are the common text, ordered text, special text, logic structure information and make up information of the document.

Main feature of building objects data is the combination of geometric solid representation and semantic project information about object. The basic elements of data structure are elements of drawing structure, geometric and non-geometric objects, space object models, building construction models, visualization data, and attributes of data items. The protocol of data interoperability must support all of these data types.

3.3 The requests to an universal data format

The following main requests to the universal format of data representation in the building projects are formulated on the ground of analysis:

- Completeness of data representation about objects and their properties; the support of the syntactic and semantic party of information;
- Possibility of use of a format on various development stages for issue, intermediate storage and perception of data;
- Possibility of the integration of new subsystems;

– Support of interactive interaction with the designers and other users;

– Operating introduction of modifications originating during designing into universal format files, without necessity of the multiple repeated requests of an additional information;

– Support of parallel work above the project in several design centers with allowance for adjusting and adoption;

– Simplicity and unambiguous of algorithms for transformation of special formats of IT subsystems into universal data format;

– Compact data representation and acceptable velocity of data reading;

– Support of interaction of IT subsystems, working in various operating systems and on various hardware platforms.

The author enters the concept "potencies of the format" as outcome of the analysis of formats for building object design systems. The potency of the format M is conditional magnitude approximately reflecting dimensionality of a transmitted data model. The systems with formats from 0 up to 5 are revealed, that generalizes a known classification of CAD models: 2D, 3D etc. It was assumed, that the universal format ($M = 5$), ensuring a storage and use of any data produced by subsystems, is a basis for infra-design of IT. Therefore the potency of an outcome can be expressed through potencies of initial formats and systems:

$$M_o = \min\{M_b, M_t, M_e\}, \tag{1}$$

Where M_b – potency of an initial format or initial system;

M_t – potency of the intermediate format;

M_e – potency of the outcome format.

The “might” of the format can also be calculated as the relation between the quantities of the right received information and the sent information:

$$M'_o = \sum_k \frac{A_{rk}}{A_{sk}} \gamma_k, \tag{2}$$

Where A_{rk} – quantity of the right received information of type “k” after transformation;

A_{sk} – quantity of the sent information of type “k”;

γ_k – the relational weight of the information of type “k”.

The methodological principles of architecture for data transfer and conversion in IT systems are developed on the ground of studies of feature of some building software. The simplified model of data transfer and conversion is showed on the fig. 1.

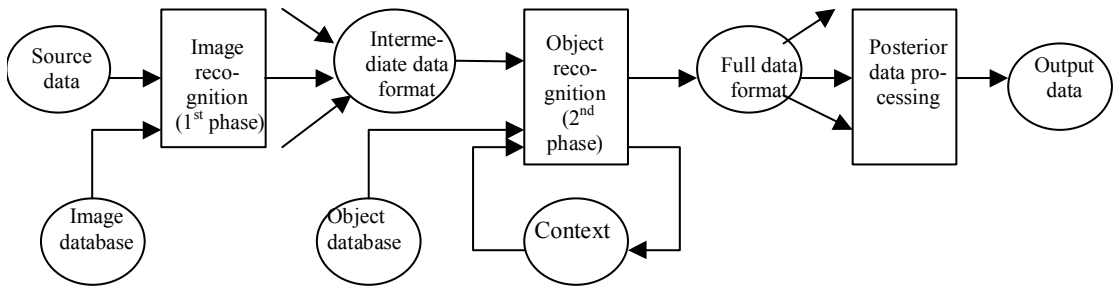

Fig. 1. The model data transfer and conversion

4 Data domain model and the data structures

4.1 Modeling of data domain for construction

It is necessary to make the formal description of construction systems and objects on the basis of modeling for the description of construction data domain. It is required also to form the classification of main data domain concepts, building systems, objects, resources, their relations, events and processes happening in data domain.

One of main points in software engineering takes the database designing. The model "entity-relationship" (ER-model) was found the wide application by information and logical modeling. However the model has some weaknesses from the point of view of construction data domain. For example, there are no special terms for such concepts as arrays, matrixes, lists etc. It is expedient to apply as attributes in dynamic building systems not only separate significances, but also function.

The comparative analysis of relational, object-oriented and hybrid databases, knowledge models (include rule-based, frame-based, logical-semantic, and mathematical models etc) was shown that the modified frame method (MFM), close to object-oriented modeling can be offered as the ground of analysis. The slots of the frames can be specialized by MFM, the amount of slots is variable, they can be inherited and to have polymorphism. Except for concept "object", there are such concepts in MFM, as "clone", "system", and "medium". The UML-language is selected for map of MFM models as an outcome of the analysis of various means of model visualization and development (IDEF4, EXPRESS-G, UML).

4.2 Structure of the building object data

The following main requests to the data domain model are offered:

- Support of information about various aspects of building object;
- Representation of outcome data of both common and special IT for building application;
- The hierarchical structure of building object types should not be strict;
- Support of the storage, transfer, transformation, recognition and addition of data;
- Representation of main properties of objects both in the data format, and in programming classes;
- Preservation of graphic, alphanumeric and binary data;
- Flexibility of an object model permitting the extension, patching, watching, and recording of modifications of information;
- Global identification of objects, witch support authorship of the project, preservation of the comments, security and confidentiality of information;
- Use of the distributed databases in a global web and support of design in the local network;
- Maintenance both late, and early data binding;
- The support of simultaneous application various hardware and system platforms.

All population of classes is divided into three main groups:

- Auxiliary classes of common purpose;
- Description of topology, geometry and visualization properties of objects;
- Classes for building objects and systems description.

4.3 UML-description of data domain

Description of data domain includes besides data types, including predetermined and changeable enumeration data. The UML-diagram of model subsystems of data domain is shown in the fig. 2.

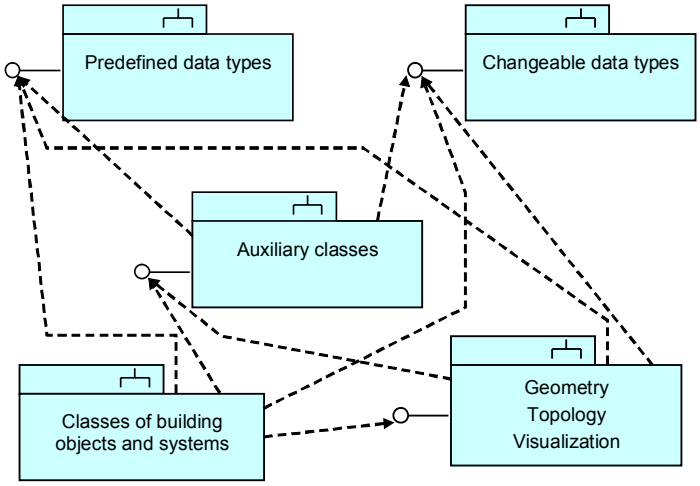

Fig. 2. The diagram of data domain model subsystems

The diagram for topological classes for solids is indicated as example in a fig. 3, one for main high-level classes – in a fig. 4 [1].

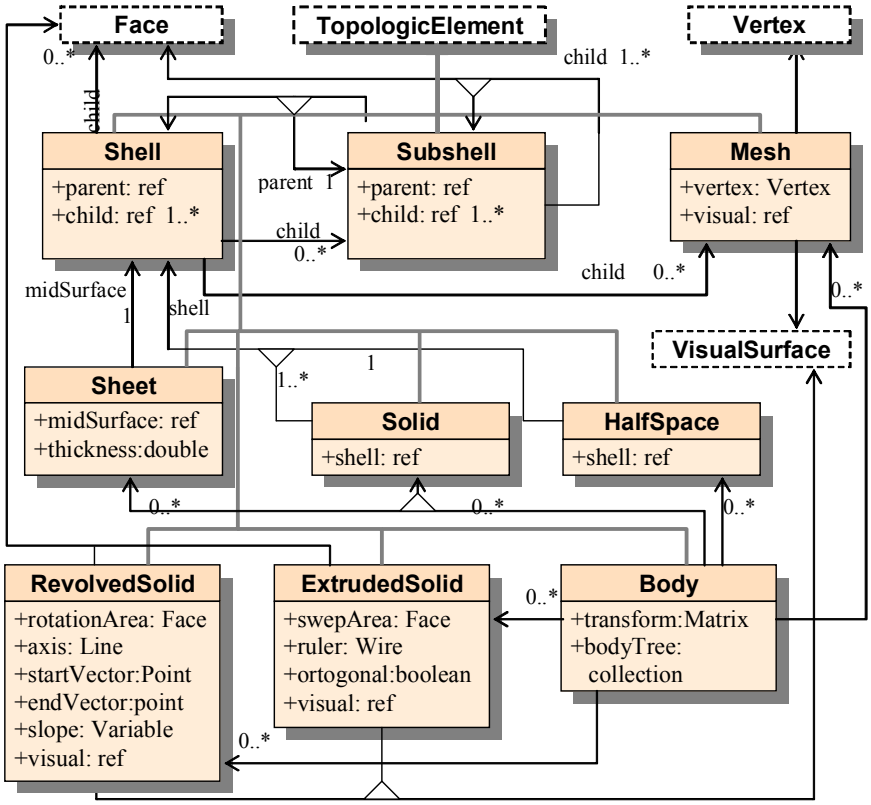

Fig. 3. The diagram of topological classes for solid representation

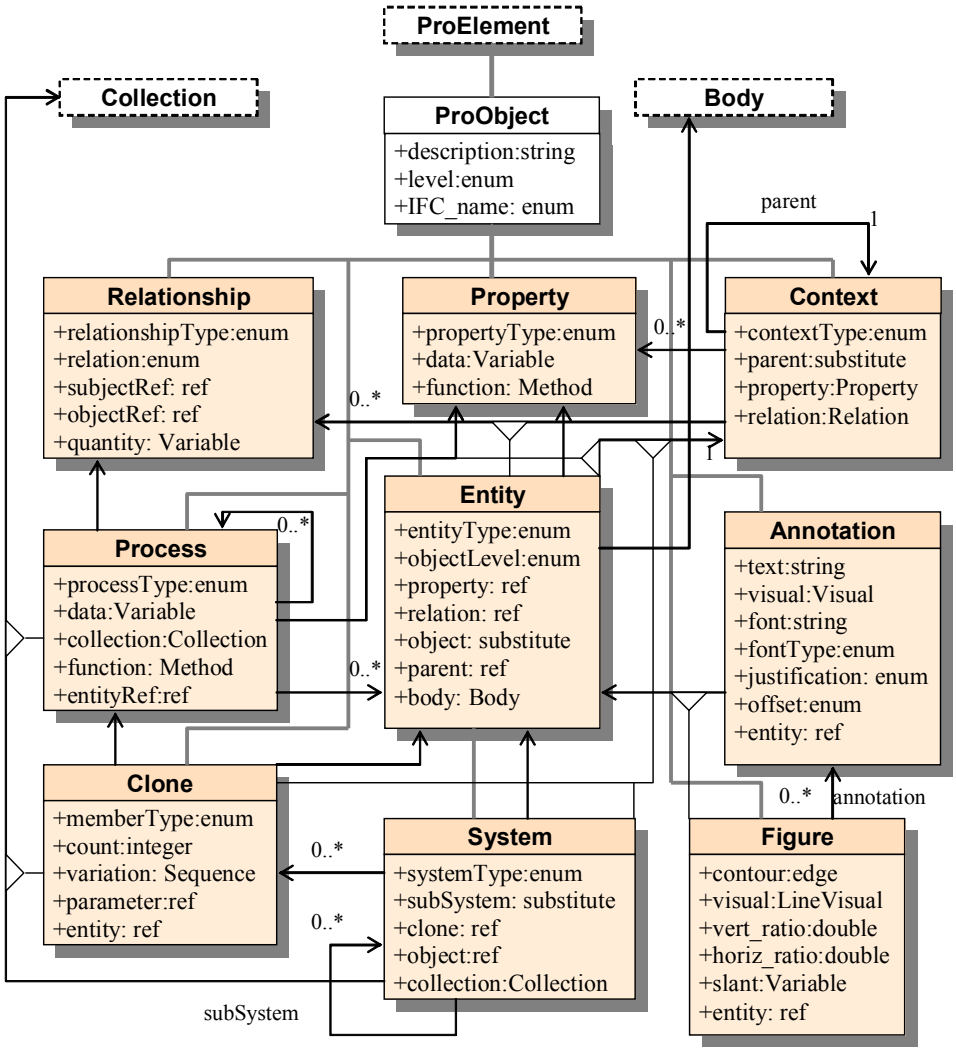

Fig. 4. The diagram of main high-level classes

The represented base set of classes allows presenting the construction data domain as an aggregate of class models of objects, systems and their relationships. The selected method of visualization ensures good layout and completeness of object representation.

5 The formalized description of building objects

5.1 Concept of the typology

On the basis of analysis of the existing theory of a classification and semantic analysis of data domain is established, that a uniform classification of technical and building objects, of their parts and properties, and also concepts used in the design documentation, practically does not exist. The author offers to name homogeneous field of object property space as "*taxon*", as the concept "class" is yet not applicable; first of all, because of use of this term in object-oriented programming as programming unit with beforehand determine structure. The following principles of typology of construction data domain are developed on the basis of indicated requests:

1. The typology of concepts should be rather *extensive* and *miscellaneous*. The following main kinds of concepts are enveloped there: types of erected objects and their parts, involved

resources, processes, conditions of both object creation and object existence, and also abstract concepts used in the descriptions, in particular, object properties and relationships.

2. It is expedient to apply a *hierarchical classification* of concepts.

3. The classified object can belong simultaneously to several taxa, enclosed each other or intersected. The multiple references of objects to taxa are allowed.

4. It is expedient to apply a logic (or deductive) classification of concepts by a descending method.

5. The classification of objects is made, as a rule, on the most essential indications, which can be consolidated in four groups of the generality: *an origin, structure, properties and application of objects*. The application of one of four property groups on each step of the branching of the hierarchical tree is determined by a concrete situation: they can alternate, combine, or one of groups can be used some branching in succession. One of the most important groups of classification indications should consider the generality of application of construction objects.

6. It is expedient to design the classification with allowance for optimization of a structure of the search tree on the criterion of velocity of interactive search.

7. The application of global identifiers of both objects, and taxa is necessary.

8. Except for a taxa hierarchy, it is necessary to develop a hierarchy of an enclosure of objects. The taxa classification should be created for each level of an enclosure of data domain objects.

So, the “Taxon” is the categories of computer modeling occupying an intermediate position between a class and object. Taxon is a class as a classification concept of the real world, however is an object from the point of view of object-oriented programming. The relationship between categories class – taxon – object is shown in the fig. 5.

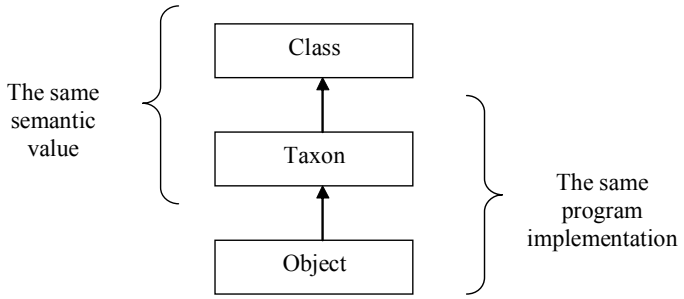

Fig. 5. The relationship between categories class – taxon – object

The written model "class – taxon – object" is implemented by the author as the specially developed prototype of object-oriented database and can be used both for study objects and taxa behavior and for practical realization of object storage in information systems.

5.2 Object coding and taxon coding

It is practicability for taxon identification to use a signifying code, on which it would be possible to reconstruct the hierarchical relationships between taxa. As such code, the combined code consisting of pairs of the description of hierarchy level and number of taxon at the given hierarchy level, for example, "**`c2.ca1.d8.da34.f3.fb14.h8`**" is offered. The combined code including both a taxon label, and label of an object enclosure level, can be showed by the following example: "**`c2.ca1.d8(HF1).f3`**". Both hierarchical coding and 128-bit numerical significance of the Universal Unique Identifier (UUID) can be applied for identification of object [2].

The offered system of coding and the being available software allows supplying unambiguous identification of taxa, objects and other elements of data domain.

5.3 Hierarchical structure of construction data domain

With the purposes of unification of hierarchical structure of main data domain concepts the universal enclosure levels of geographical objects, objects of town-planning, architecture, design, and construction objects are developed [3]. For both support of object recognition by it properties and definitions of its cost and technological parameters it is necessary for formalizing properties of building object. The groups of building objects properties are chosen on the basis of studies of a number of the construction norms and technology documents. Those are structural, topological, geometric, constructive, production, operational, technological, economic, legal, and also property of a material and object surface. The said properties can be formulated by constants, physical regularities, behavioral responses of object, production functions, text description etc.

For the formal description of data domain is made besides the classification of the relationships between objects and taxa, their postures, processes and resources of construction is executed. Thus, it is possible to select as the most important among the *relationships* the structure, cause and effect, origin, management, comparison etc. It is possible to refer to main properties of building objects *conditions* the spatial-temporal position, physical, chemical, production and technological condition, and also legal status. Main kinds of *processes*, happening in building objects, are transition, modification of the physical or chemical condition, the production and technological processes, modification of the legal status etc. The main *resources* of construction are information, financial, labor, material, organizational, and technological resources.

5.4 Information context of building object

For consideration of building object in correlation with other objects or systems the author enters concept of an information context (*IC*) of an object E_i: the *context* is a fragment of the information medium, which consideration is necessary for relevant study of object in definite purposes. The information medium *IM* of a designed system *S* is a set of the relationships *R* $\{r_{ij} | <E_i, E_j>, E \in S\}$, in which the various objects *E* can (or could) be with other objects or systems during life cycle.

Therefore a system context is association of contexts of subsystems and objects, which are included in a system:

$$IC_s = IC_1 \cap IC_2 \cap \ldots IC_i \cap \ldots \cap IC_n, \quad i \in S . \tag{3}$$

The context IC_s of all designed system is a global context in relation to contexts of included subsystems and objects and is, as is said above, information medium of a system. The context describes conditions, in which exist objects. The purport of separate a context from the object that they can be changed independent from each other. It ensures safety of data entered in the interactive dialogue and take off the necessity to repeat the requests during continuous transfer and updating of data.

The context can have the following elements: a code, reference on superior context, spatial-temporal restrictions, parameters etc. The contexts can form the hierarchy: global context, intermediate contexts and local context. The code of a context consists of an object code, to which it concerns, with adding of a sub-string "**.ctx**". The groups of the contexts (social, economic, normative, legal, geophysical, ecological, architectural, town-planning, constructive, management, technological, operational context etc), are allot on the basis of consideration of designing of civil, industrial and power buildings and structures.

6 Data transfer and object recognition

6.1 XML-language for description of construction objects

The author has set up the problem of creation of universal data transfer format for building objects for significant reduction of document volume and acceleration of processing time. Main kinds of useful databases and formats (dBase, Paradox, DXF, SAT, GDL, STEP, SGML etc.) were consideration. The XML-language is selected as the base of the universal data format, ensuring natural hierarchy of objects, flexibility, good layout and expandability. For reaching the goal, it was necessary also to decide the following tasks:

– Creation the document scheme for the data domain description;

– Minimization of class variety;

– Supply a strict outline of the scheme permitting to map it easy with the help of various programming languages;

– Developing language structures support a repeat of identical objects and a reduction of transmitted information;

– Distributing of information between the various schemes allows simplifying a procedure of their modification.

The language, developed by the author, is called Building Object Description Extensible Markup Language, or bodXML [3]. The scheme of language is divided into the description of separate elements representing various data types of the document (fig. 6). Language completely bases on the concept of modeling of construction data domain writing in a chapter 4. In the scheme are described:

- Low-level elements (variables, references, arrays etc.);
- Specialized elements (geometric, topological etc.);
- High-level elements (objects, their population, system).

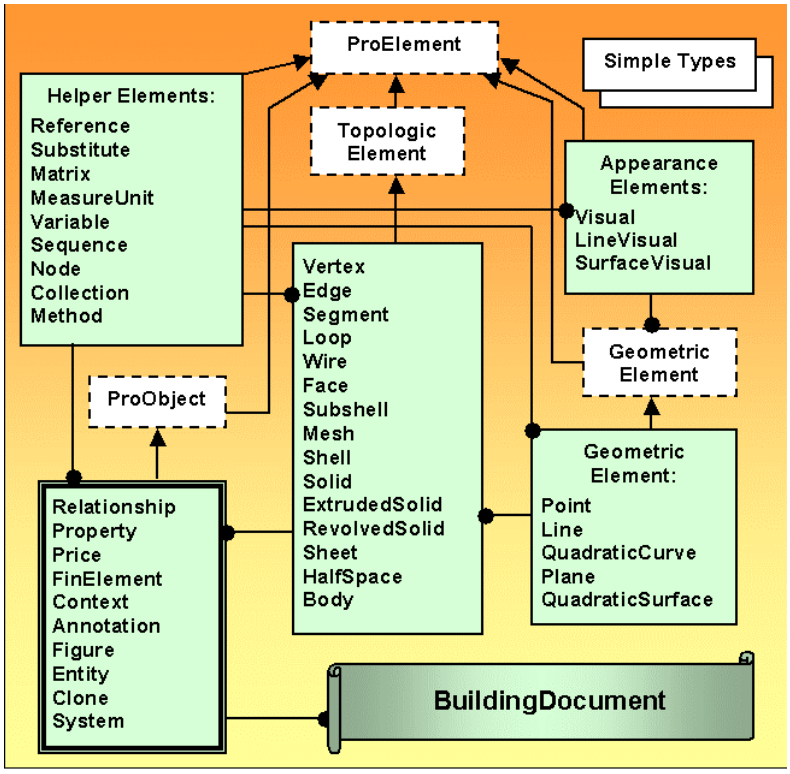

Fig. 6. The diagram of inheriting of scheme elements

By development of language was changed the traditional approach to parity between the schemes and documents: the use of databases composed as the XML documents was entered. The scheme includes a rather small set of elements (about 50), essentially distinguishing on a structure and necessary for brief but detailed description of the documents. The significant part of these elements is entered for the adequate description of the geometric form of building constructions.

6.2 Use of the distributed databases

The author offers to place in the global computer web not only the scheme of language but the common databases describing widely widespread concepts, for example, classification of building objects and resources, unified descriptions of building materials. The lists of the most important objects, for example, units of measurements, line styles and textures of surfaces, geographical and toponymic objects must be disposing in global web. These databases should be supported by the manufacturers and users of the software associated to building design. The mutually agreed modifications without scheme modification can be introduced to the distributed databases. It ensures stability of the software versions. The prototypes of some databases for construction are developed by the author and are showed on his Internet home page [3]. The examples of the objects description of a various level with the help of universal format bodXML are indicated there too.

6.3 Object recognition by the data transmission

One of key moment for provision of interaction of various information systems is the recognition of transmitted objects. In the accepting system is thereby a problem of creation of objects of such classes, which were not provided by development of the given system. The modern means of object-oriented programming do not support a possibility of dynamic creation of new classes.

To bypass this restriction, essential to a considered problem, the author offers (chapter 5) to enter the concept “taxon” into consideration. Distinctive feature of the dual nature of taxon is the availability of properties incident to both class, and object. The objects of one taxon can be program identical to objects of other taxon with the similar structure of properties. Therefore the new taxon can be created at the run-time instead of compilation, other words, without the modification of the program. It is “quasi-dynamic” run-time creation of the new class.

Main feature of building objects data is the combination of geometric solid representation and semantic information about object. The analysis of data structure for such objects has shown that the data record of object or taxon can include three main parts: service, specific and semantic part. The geometric solid of objects can be represented by various methods: as dot objects without the form description, elongate body with the description of an axial line and section, flat body with the medial surface and thickness, 3D-solid with boundary representation.

The standardized name of a layer can be used for preliminary recognition of object type. The international standard ISO 13567 can be applied to this effect. For the more detailed semantic analysis of object the application of artificial neuron network algorithms is possible.

The author offers a set of main topological and geometric properties being sufficient for recognition of main three-dimensional building constructions with flat edges. Such properties are, in particular, ratio of overall dimensions along main orthogonal axes, ratio of valid and overall volume, asymmetry of vertices coordinates, enclosure of face outlines, direction and character of the enclosed load etc. The tests of artificial neuron network have shown, that the recognition of a kind of the constructions represented as a set of indicated parameters happens enough confidently.

7 Endnotes

The creation of integrated IT, predetermined by technical progress in construction, requires the interoperability of heterogeneous software. Without this the efficiency of software application in design firms is reduced. It confirms that the main purpose is the creation of scientific and methodical bases of data transfer between IT in construction.

The universal structure of the construction data domain description permitting to establish the scientific fundamentals of the semantic accurate data transfer and interoperability of information systems is created with the help of UML-diagrams.

The data structure for object recognition ensuring quasi-dynamic creation of classes during data transformation is developed. The concept "taxon", occupying an intermediate position between "class" and "object" is entered for this purpose. Taxon is a class from the point of view of classification of the real world objects; however it is an object from the point of view of object-oriented programming.

The concept "design information context" is entered and their structure is developed permitting to facilitate essential recognition of building objects during data transfer. The main types of contexts are chosen: social, economic, normatively, legal, geophysical, ecological, architectural, town-planning, constructive, management, technological, operational etc.

The scientific and methodical fundamentals of the formalized description of construction data domain are created with the help of developed the language "bodXML" that is a dialect of XML-language. It permits to represent main features of building objects and to transmit significant information about them.

The technique of building objects recognition represented as 3D-solid models, with the help of artificial neuron networks is created. The list of properties of objects permitting with the sufficient degree of reliability is developed to recognize main elements of building constructions.

8 References

1. Pavlov A. Die Modellierung des Bauwesens für den Datenaustausch // IKM-2003, Bauhaus-Universität Weimar.
2. Pavlov A.S. The Language for Description of Construction Data Domain./ XI Polish-Russian seminar "Theoretical fundamentals of construction". Moscow-Warsaw, 2002.
3. http://www.mtu-net.ru/pavlov/eng