Classification and Coding of Entities of Construction Data Domain

A.S. Pavlov
Moscow State University of Civil Engineering, Russia

D.S. Kulitchkov
Russian President Academy for the civil servants,
Institute of advanced training for the civil servants, Moscow, Russia

ABSTRACT: The classification of objects and their properties is necessary for the semantic description of construction data domain. The authors have offered the facet-hierarchical classification of entities in construction. The authors propose also the project of classification of main concepts in data domain of construction: settlements, territories, buildings, resources, properties etc.

1 INTRODUCTION

The various software are widely applied at present in practice of building design. The software have formed by the different development engineers and uses various databases and formats for a data storage and data transfer. Thereby the problem of interoperability appears, in another word that is compatibility of various software products among themselves.

The interoperability of information systems is divided on syntactical and semantic parts of object-oriented description. The syntax abstracts both from concrete objects, and from their sense. It is applicable for the formal description of the object's properties. Semantic component is the description of a sense of concepts in generalized classes of data domain objects. The semantic analysis of data domain is therefore the development of a hierarchical classification of main concepts used in construction. Such classification, though is not the comprehensive semantic description, ensures improving interaction of information systems in the construction.

2 SEMANTICS AND CLASSIFICATION

Gottlob Frege (1848-1925) and Ferdinand de Saussure (1857-1913) have laid the main principles of a linguistic semiotics. The founders of philosophical semantics were Alfred Korzybski (1879-1950) and Rudolph Carnap (1891-1970). The "constructions theory" of Carnap consist that some concepts express analytically through other concepts, and forms the hierarchy. Empirical and axiomatic concepts are the basis of a hierarchy. The artificial concepts can form the high levels of hierarchy.

The dual notion of any classification system is necessary for the complete description: that are taxonomic (i.e. typology of objects) and meronomic (i.e. formalized description of properties).

A point in the **n**-dimensional space of its attributes can represent any object. To classify met object it means to refer it to one of being available groups of objects. Coordinates of objects were in this area, are the attributes of object (merons).

Many classification systems are used, namely enumerated, hierarchical, facet, many-dimensional, etc. The hierarchical and facet classification methods are used most frequently.

The subsets are strictly enclosed each other by hierarchical method. The hierarchical method allows us to construct a concept tree of any depth that enables to classify large manifold of information. This method requires however the beforehand ordering of the attribute set and does not allow the including of new attribute in a middle of the tree structure.

The facet classification is based on parallel separation of generally set of objects on the independent classification attributes. The facet method is more flexible, and it allows the look-up on the arbitrary combination of attributes. The length of a code is relatively big, but this defect cannot be considered essential in conditions of the up-to-date information processing. The absence of a hierarchical system of interactive access to the complete concept list is more serious defect.

Most convenient is the combined hierarchical-facet classification. This principle is used for example in Statistical classification of economic activities in the European Community.

3 CLASSIFICATION IN BUILDING ACIVITY

The qualifiers in data domain of construction are based mainly on the functional attributes, i.e. on the purpose of objects. The basis of the classification of production and services in Russia are the International Standard Industrial Classification of all Economic Activities - ISIC and Central Product Classification - CPC. The code of branch is taken four digits and the code of product – three digits. The classification comprises more than 4200 concepts relating construction, including more than 1000 kinds of works and services, more than 1200 names of building constructions, details and other resources, more than 1700 kinds of the enterprises, buildings and structures, about other 300 abstract concepts (sections of the projects, performances of buildings etc.).

The Russian qualifier of production comprises around 52500 groups and names of production, including more than 5 thousand names of material resources used in construction (not counting the productive facilities). The amount of kinds of resources used in construction is significant more. The Russian state estimate code is contained about 28 thousand names and marks of materials and construction item.

The classification of objects will actualize the paradigm of object formation "ancestor – descendant". But it is necessary to describe such relations, as "consists from…" and "has properties…" and other. Really, the object of construction is characterized by the internal structure, properties etc.

The authors consider, that the system of concept and objects classification for construction data domain should have the following properties:

– The amount of concepts must comprehend the really and abstract concepts of business processes not only for building, but also for adjacent data domain

– The structure of classification must have the internal logic; that will permit fast search both in interactive or automatic mode

– The classification must comprises not only concepts, but also properties of objects

– The classification must describe the enclosure of objects

– The depth and breadth of a branching should be adhering to optimum, to be sufficient for accommodation of a fair quantity (up to hundreds thousands) of concepts

– The addition and modifications of a classification structure at various levels should be flexible without a modification before the created documents

– The additions in the distributed databases must be possible

– The actual information technology must be applicable.

4 THE PRINCIPLES OF CLASSIFICATION

The authors are developed the following principles of classification in construction data domain for sufficing the indicated requests.

1. The construction business is the large open dynamic system. The typology of concepts of such data domain should be rather extensive and multiform. The following main kinds of concepts should be enveloped: types of building objects and their parts, resources, processes both conditions of creation and existence of objects, property and relationship of objects, abstract concepts used in the descriptions. By object-oriented modeling it is necessary to create classes simulating objects and concepts. The attachment of objects to concept is carried out by creation of the reference on the taxon (an element simulating concept), or by the global identifier.

2. Is proved, that any structure of data domain can be shown as hierarchical classification. That allows the unambiguous identification of concept. Thus, it is useful to apply a facet-hierarchical classification of concepts. The concept hierarchy in computer models can be supplied by the reference on a higher level of a classification. In artificial systems (the classification of an information, production etc.) it is usually enough 5-10 levels of a hierarchy, in natural systems (for example, biological) - up to 20 and more levels.

3. The existing systems of classification are represented as the tree graph or forest graph not have cycles and intersections. The real objects (both natural, and artificial) are not always interpretable as strict classifications. For example, the waterproof can be classifier as roof systems or fundament systems; the steel pipes can be used as constructive bearing elements, for HVAC system or for installation of the productive facilities etc. The object can therefore belong simultaneously to several concepts, enclosed each other or intersected - such method is named as a poly-classification. The multiple references on concept suppose more high levels in other branches for realization of this method.

4. The construction data domain comprehends the objects both artificial, and natural origin. Artificial objects play the more significant role. By classifier of artificial objects the inductive method cannot be guaranteed logical convergence of a hierarchy in the upper point. It is expedient to apply a deductive top-down classification of concepts. At the uppermost level can be chosen the concepts, which are divided in turn into natural objects and artificial objects.

5. The history of development of a science shows, that each science creates the typology of investigated objects in own way. The classification of objects however is made on their most essential indications, which can be composed in four groups of a generality: an origin, structure, properties and purpose of objects. So, the typology of biological or geological

objects is based on the indications of a common origin, typology of chemical substances – on their structure, typology of physical bodies – on their properties and so one. For artificial objects a basis of a classification frequently becomes area of their purpose. The application of one of four groups at each step of the branching of the hierarchical tree is defined by the concrete situation: they can alternate, combine, or repeat.

Many classifications can exist because every material object has several various attributes. For correct perception of information it is expedient to select the main classification, which should be demonstrative, argumentative and should allow addition and modification. As shown is higher, the revealing of concepts is not univalent. It is possible to choice only one of variants of typology solution, par example on cause of usability. One of the most important attributes groups is a generality of application of construction objects.

6. Hierarchical classification is necessary for practical organization of interactive man-machine searching and entering of information, because it is enough for computer systems to ensure uniqueness of a code. That is simply reached by serial encoding of concepts. The significant time is necessary for looking up the concept by the interactive dialogue with the computer. The classification should be designed therefore so that the search of the necessary element would take short time. The structure of the search tree must be optimized for this purpose with allowance of properties of man-machine dialogue. Then the optimum branching of a tree should be defined by the criterion of a velocity of interactive search.

7. Both taxons and objects must have the global identifiers for univalent recognition of objects. The construction of the taxon identifier should not limit possibilities of extension the intermediate hierarchy levels. The combined code with reflecting of a level of a hierarchy, and taxon position at the given level is possible.

8. Nobody has been yet create a uniform classification of all concepts. Not only concept, but also objects have a hierarchical nature: they are enclosed each other and by that they will make a hierarchy system. To see analogies of wildlife, the biosphere of a planet is constructed on a principle: biosphere - population - organism - functional system - organ - tissue - cell. Each object is nested in another object of a highest level, for example, the organism is not a part of other organism, but is a part of a population. Every level of the classification is not conterminous with another level. That can be seen in microcosm (molecule - atom - nucleus - particle) and macrocosm (metagalaxy - galaxy – star conglomeration - heavenly body).

5 CLASSIFICATION OF CONCEPTS

It is necessary to develop a hierarchy of concepts in accordance with hierarchy of nested objects. In construction data domain it means about nested building systems. The classification of concepts should be created for each nested objects level of data domain. The association of such classifications in certain "super-classification" has not real sense, however it can be used for computer processing of encoding. The principal scheme of classifications of data domain is shown in a Figure 1.

Fig. 1. The scheme of classifications on the data domain at the various levels of nested objects

It is expedient to use a signifying encoding for identification of concept, on which it would be possible, the reconstruction of hierarchical relations between concepts. The combined encoding is offered. It consists of pairs of the hierarchy level description and concept number at the given hierarchy level. We shall designate a hierarchy level by the lower case Latin letter. The second letter designates second sublevel etc. Decimal digits following the level letters designate the number of concept. The amount of sublevels and numbers is not limited. Points as facet encoding of qualifiers divide the pairs. At each level there can be some classifications (for example, classification on various attributes of objects). They are numbered within the limits of a level. Let's note a concept code in the EBNF (Extended Backus-Naur Format, ISO/IEC 14977):

```
lower ::= 'a'|'b'|'c'|'d'|'e'|'f'|'g'|'h'|'i'|
          'j'|'k'|'l'|'m'|'n'|'o'|'p'|'q'|'r'|
          's'|'t'|'u'|'v'|'w'|'x'|'y'|'z';
  level  ::= {lower}+ ;
  pair   ::= level number ;
  concept_code  ::= pair {"." pair }* ;
```

Example: «c2.ca1.d8.da34.f3.fb14.h8».

Such encoding ensures arbitrary escalating and modification of a classification structure during realization. The restriction concerns only adding of highest levels: there are no letters up to the letter "a", therefore it is necessary to enter reserve highest

levels, par example, levels "a" and "b". Only declaration order of registration is necessary for addition the new concepts in a global web with minimum centralization. The scheme of a hierarchical classification with application of a combined encoding is shown in a Figure 2. The dotted lines are shown the additional references to other branch. It is enough for creation the hierarchy to have the references to higher hierarchy levels.

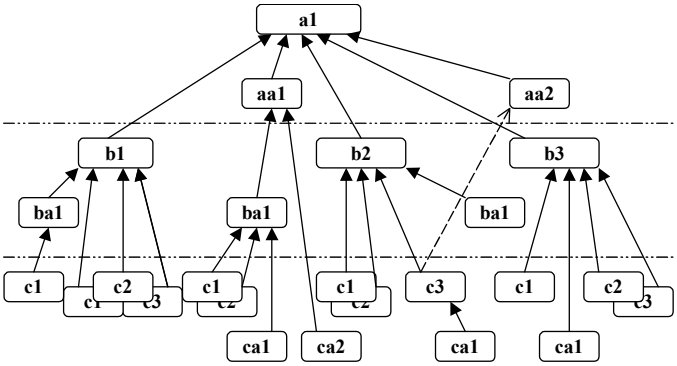

Fig. 2. The scheme of the hierarchical classification

Let's establish as a first approximation hierarchy of data domains to reserve the area of encoding for data domain of construction. The fragment of classification of buildings on destination is shown in Table 1.

Table 1.Top level of data domains

Code	Data domains
b1	Abstract entities
b2	Nature
b3	Life
b4	Noetic entities
b5	Industry
b5.c1	Traditional production
b5.c2	Mining
b5.c3	Energetic
b5.c4	Food processing industry
b5.c5	Textile industry
b5.c6	Manufacturing industry
b5.c7	Chemical technology
b5.c8	Metallurgy
b5.c9	Mechanical engineering
b5.c10	Electronics engineering
b5.c11	Building
b5.c11.d1	Building objects
b5.c11.d2	Design
b5.c11.d3	Resources
b5.c11.d4	Technology
b5.c11.d5	Facility management
b5.c11.d6	Hydraulic engineering
b5.c12	Transportation
b5.c13	Trade
…	etc.

The authors develop the classifications of some main concepts in construction data domain: settlements, territories, resources, properties etc. The fragment of classification of buildings on destination is shown in Table 2.

Table 2.Classification of buildings

Code	Type
b5.c11.d1.eh1…	Building objects:
...f1	Residential building
...f1.fa1	Capital residential building
...f1.fa1.g1	One-apartment residential building
...f1.fa1.g2	Multi-apartment residential building
...f1.fa1.g2.ga1	Multi-apartment building up to 2 floors
...f1.fa1.g2.ga2	Multi-apartment building up to 5 floors
...f1.fa1.g2.ga3	Multi-apartment building up to 9 floors
...f1.fa1.g3	Hostel
...f1.fa1.g4	School sleeping house
...f1.fa1.g5	Sleeping house for disabled persons
...f1.fa1.g6	Sleeping house of sanatorium
...f1.fa2	Non-capital resident building
...f1.fa2.g1	Mobile camping-van
...f1.fa2.g2	Container-house
...f1.fa2.g3	Garden house
...f2	Public office building
...f2.fa1	Sales and catering building
...f2.fa1.g1	Shop
...f2.fa1.g2	Trade pavilion
...f2.fa1.g3	Emporium
...f2.fa1.g4	Apothecary's
...f2.fa1.g5	Restaurant
...f2.fa1.g6	Coffeehouse, Teashop
...f2.fa1.g7	Canteen
...f2.fa2	Public utilities
...f2.fa2.g1	Hotel
...f2.fa2.g2	Motel
...f2.fa2.g3	Public service, laundry
...f2.fa2.g4	Automobile service
...f2.fa2.g5	Garage
...f2.fa3	Health services building
...f2.fa3.g1	Clinic
...f2.fa3.g2	Out-patient hospital
...f2.fa3.g3	Day hospital
...f2.fa3.g4	Health center
...f2.fa3.g5	Maternity hospital
...f2.fa4	Education building
...f2.fa4.g1	Nursery school, child center
...f2.fa4.g2	Comprehensive school
...f2.fa4.g3	Art school
...f2.fa4.g4	High school
...f2.fa4.g5	Retraining school
...f2.fa5	Culture building
...f2.fa5.g1	Museum
...f2.fa5.g2	Library
...f2.fa5.g3	Archive
...f2.fa5.g4	Studio
...f2.fa5.g5	Exhibition hall
...f2.fa5.g6	Philharmonic
...f2.fa5.g7	Theatre
...f2.fa5.g8	Cinema
...f2.fa5.g9	Concert hall
...f2.fa5.g10	Studio
...f2.fa5.g11	Sport hall
...f2.fa5.g12	Planetarium
...f2.fa5.g13	Dancing
...f2.fa5.g14	Culture center
...f2.fa5.g15	Club
...f2.fa5.g16	Circus
	etc.

The separate part classification is made for design documentation and technical rules. Example of the concept code for documentation: "b5.c11.d2.e2.f3.ff1.g2.gf8.gk1.hg8.hk3.hp2.hu5".

It means "The territorial quotations for building rehabilitation".

6 CLASSIFICATION OF OBJECTS

The 128-bit numerical significance of the universal unique identifier UUID (or GUID) is applied for identification an object in memory of the computer. In this significance first 60 bits are used for an entry of discrete time of object formation (amount of 100-nanosecond intervals from introduction of Gregorian calendar style). The 14 bits are used for the computer clock synchronization sequence and 48 bits – for address of computer knot in the global web. Remaining bits are used for version number and attributes field. The standard function `CoCreateGuid (GUID*)` can be used for obtaining UUID in Microsoft environment space. The casual digits are generated, if any of necessary data are inaccessible. UUID in 16-digit format with separators takes 36 bytes, for example: "`AC5C6360-248C-447F-8976-92A3F102 9507`".

The uniqueness of a model of object in memory of the computer however does not guarantee against a collision by creation of several models of the same real object. It can result in dissociation of data about object and to distortion in the design documentation. It is expedient to appropriate a unique identifier not to model but to object itself. The offered scheme of encoding of nested objects can be used for this purpose. The unique encoding of construction should include codes of administrative or territorial division, town-planning and architectural documentation. Par example, the encoding of construction located in Moscow, can look as follows: «`CP643.DF46.DP280.DU597.EP1.FK12.FP0099`». The similar method can be used for the constructions, elements of buildings and so one.

The level of nested objects is a basis of hierarchy division of data domain. The preliminary study of the structure of nested construction objects has shown, that for the description of such structures there are enough 12-16 basic levels. For a label of such levels we offer to use capitals of the Latin alphabet. The sublevels etc can be added if necessary. Then in the EBNF the entry of encode of a □lassifycation can look as follows:

```
digit::='0'|'1'|'2'|'3'|'4'|'5'|'6'|'7'|'8'|'9';
upper::='A'|'B'|'C'|'D'|'E'|'F'|'G'|'H'|'I'|
        'J'|'K'|'L'|'M'|'N'|'O'|'P'|'Q'|'R'|
        'S'|'T'|'U'|'V'|'W'|'X'|'Y'|'Z';
nest_level   ::= {upper}+ ;
number       ::= {digit}+ ;
classif_code ::= nest_level number ;
```

Par example: «`DA2`».

Let's enter the generalized uniform conditional names of nested hierarchy levels for all data domains (Tab. 3). Any additional sublevels can be added of course. The common abstract concepts are not included into this system, as are not objects.

Table 3. Generic and building nesting object levels

Level code	Generic level	Building object level
A	Generic	
B	Concept	
C	Global	Geography
D	Common	Regional
E	Major	Urbanistic
F	Grand	Architectural
G	Prevalent	Building
H	Extended	Room
I	Individual	Construction
J	Special	Industrial
K	Specific	Technical
L	Distinct	Production
M	Partial	
N	Single	
O	Chemical	
P	Physical	

The levels A and B are reserved for the description of microcosm, and lowest levels – for microcosm. They will coincide for any data domains. Level "I" at center of a level structure is supposed for the man, as "measures of all things".

The levels C and D are expedient for fixing behind geographical concepts, common for many data domains with real "terrestrial" objects. The encoding of sublevels has reservation of codes (between named sublevels four reserve unnamed sublevels).

The fragment of nested building object sublevels is shown in Table 4.

Table 4. Hierarchy of construction object sublevels

Sublevel	Definitions and examples
FF. Ensemble	Element of zone, limited to arterial roads and other boundaries (ensemble, group of quarters or enterprises)
FK. Quarter	Element of buildings limited to streets (microregion, quarter, block of enterprises)
FP. Site	The group of objects, which building construction is carried out on a one project
FU. Locality	Part of building site, were in one place
GF. Stage	The project determined part of the building construction, created without interruption
GK. Start-Up	The part of stage, ensuring output of production or services (power plant unit etc)
GP. Group	The technologically completed part of a complex (facilities, some homogeneous constructions)
GU. Subgroup	The homogeneous part of a group, technologically unit
HF. Building	Isolated object of the immovable (building, structure, road, planting)
HK. Section	The part of building (Entrance, block, workshop, technologic train)
HP. Space	Isolated part of section (floor, cellar, aisle)

Table 4. Continued

HU. Room	Isolated part of space (room, staircase, bay)
IF. Structure	The functionally homogeneous part of a building (foundation basis, foundation system, bridging, walls, roof, HVAC system)
IK. Block	The spaced or constructional homogeneous part of a structure (foundation, wall, ceiling, roof batter, part of HVAC system)
IP. Facility	Mono-functional part of the block (unit of equipment)
IU. Construction	Separate or repeating construction (opening, plate, panel, column, beam, farm, down-comer)
JF. Aggregate	Assembled part of a construction (reinforcement, door, window, pipeline, aggregate)
JK. Device	The composite device (3D bar-type reinforcement skeleton)
JP. Part	The part of device (console of column, bar reinforcement mat, flat skeleton)
JU. Component	The component of construction (fabric reinforcement mesh, parallel-lay-wire tendon)
KF. Unit	Assembled component of a construction (unit, shutter, support, pipeline, cable, reinforcement position: rebar, array, mesh)
KK. Subunit	The constituent of a unit (furniture, valve)
KP. Pack	Gang of homogeneous elements (mark of reinforcement)
KU. Element	Element of construction
LF. Item	Accessory item (longitudinal reinforcement bundle, embedded item, stretching wire)
LK. Subitem	Part of item (gusset, corner plate, hitch plate, bolt)
LP. Detail	One-piece element of a construction (single rebar, binder, blanket, layer)
LU. Subdetail	The least homogeneous part of a details (edge, seam, hole)

7 CLASSIFICATION OF PROCESSES

The life cycle of production is the complex of processes from initial revealing of market needs before final sufficing of the raised requests. According to standard [ISO 9004-1-94], the main stages of life cycle are: marketing, development of production, process planning, purchase, production, check, packing and storage, realization and distribution, mounting and commissioning, service, after-sale activity, recycling.

The aforesaid list of stages is typical just for large-scale industrial production. For construction business-processes, which differs by uniqueness and by large duration of objects creation, not process, but project approach is incident. The concept of life cycle for project management is formulated in [Pavlov 2003]. Life cycle of building object is practically identical with the investment cycle. The following list of stages of building object life cycle is offered:

- Creation of an intention of the project;
- Substantiation and planning of the investments;
- Research and experimental effort;
- Exploration and design;
- Realization of contract tenders;
- Contract conclusion;
- Preparation of building production;
- Erection of buildings and structures;
- Mounting of the equipment;
- Laying of the communications;
- Starting-up and adjustment works;
- Going into operation;
- Assimilation of a potency;
- Realization of trial production consignments;
- Operation;
- Repairs;
- Renovation;
- Termination of operation;
- Liquidation.

8 CONCLUSION

The classification of objects and their properties is necessary for the semantic description of construction data domain. The authors have offered the facet-hierarchical classification of entities in construction. Many classifications exist simultaneously at different levels, for example, classification of buildings, constructions, and projects. It is necessary to divide first of all such levels from each other. At parallel levels are located the independent entities: objects, resources, properties, processes etc.

The codification of entities is necessary for an entry in an electronic and paper kind. The Latin letters designates each level. The concept of one level is differed by number. The letters and the numbers will derivate a unique code of concept. The intermediate levels can be added arbitrary.

The individual object has the unique identifier and code of their entity. The object has also properties, which have the value and code of entity. All objects thus differ and can be circumscribed with a desirable degree of a detail. The authors propose the project of classification of main concepts in data domain of construction: settlements, territories, buildings, resources, etc. The top levels of a classification of building objects properties, classifications of the relations, conditions, processes and resources are developed also.

REFERENCES

ISO 9004-1-94. *Quality management and quality system elements*. Part 1. Guidelines

Pavlov A. *Information transfers and object recognition in CAD-systems of building design*. M.: Millennium Foundation, 2003.