IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Constructive concept of address normalization

To cite this article: G M Rudakova *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **537** 052007

IOP Conf. Series: Materials Science and Engineering **537** (2019) 052007 doi:10.1088/1757-899X/537/5/052007

Constructive concept of address normalization

G M Rudakova[1], S L Gladkov[2] and O V Korchevskaya[1]

[1] Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation
[2] LLC Aigeo, 105, office 4-34, 60 years of October ul., Krasnoyarsk, 660079, Russian Federation

E-mail: gmrfait@gmail.com

Abstract. The paper addresses the problem of addressing as an identifier for the location of real estate objects. One of the main tasks in this field of activity is the problem of reducing an arbitrary address to normal form. From the analysis of the definitions of the address it follows that the address is a point in the space of the requisites. The constructive concept of the normalization of the address and requisites is introduced. The set of normal addresses (requisites) refers to the union of the base and revoked (canceled) addresses (requisites). The real set of values of each requisite is wider than the set of its normal values. Therefore, the steps of expanding the set of normal values of the properties in the direction of the set of real values of properties are considered. At each step of the proposed method, the rule (function) of normalization of the extended requisite value is considered, i. e. converting it to one of the normal values. The requirements for the synthesis of the properties of the requisite in the form of a name and number, spelling, following the words and the uniqueness of its semantic meaning, as well as to the semantic normalization of addresses are given.

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. Sets of normal addresses and requisites

The set of normal addresses will be called the union of the base and canceled addresses:

$$\overline{A} \equiv A_b \cup A_c; \quad (1)$$

IOP Conf. Series: Materials Science and Engineering **537** (2019) 052007 doi:10.1088/1757-899X/537/5/052007

where $A_b \equiv A$ is set of base addresses, possessing properties of uniqueness, obligatoriness of legitimacy, relevance and unambiguity, A_c is set of revoked (canceled) addresses [4]. Sets of base and revoked addresses do not overlap $A_b \cap A_c = \varnothing$ [5].

F – the function one-to-one displays a set of base addresses to a set of addressing objects:

$$A_b \overset{F}{\leftrightarrow} O_b . \tag{2}$$

As noted earlier, the set of addresses is a subset of the Cartesian product of sets of values of requisites $A \subset R_1 \times R_2 \times \ldots \times R_n$. Therefore, the set of normal addresses belongs to the Cartesian product of the sets of normal values of requisites.

R_{bi} is called the basic subset of all values of the requisite R_i $i \in [1, n]$ for which it is true:

$$\forall r_{bi} \in R_{bi} \ \exists \ a_b \in A_b \wedge a_b \in R_1 \times R_2 \times \ldots \times R_{i-1} \times r_{bi} \times R_{i+1} \times \ldots \times R_n . \tag{3}$$

The set of base values for each requisite is the value of requisites that are present in the set of base addresses.

R_{ci} is called a subset of the revoked values of requisite R_i $i \in [1 ,n]$ if

$$\forall r_{ci} \in R_{ci} \longrightarrow \forall r_{ci} \notin R_{bi} \wedge$$
$$\exists \ a_c \in A_c \wedge a_c \wedge R_1 \times R_2 \times \ldots \times R_{i-1} \times r_{ci} \times R_{i+1} \times \ldots \times R_n . \tag{4}$$

The set of revoked values of each requisite is the value of requisites that are present in the set of revoked addresses, but are missing in the set of basic addresses of the requisite.

Thus, the set of normal properties of a requisite is called the union of the subsets of the base and revoked values of the requisite:

$$\overline{R} \equiv R_b \cup R_c \tag{5}$$

From the definition of the set of normal values of the requisite, it follows that the intersection of the sets of basic and revoked values of each requisite is empty:

$$R_b \cap R_c = \varnothing . \tag{6}$$

This property follows from the definition of the set of revoked requisites values.

The set of base addresses belongs to the Cartesian product of sets of basic values of requisites:

$$A_b \subseteq R_{b1} \times R_{b2} \times \ldots \times R_{bi} \times \ldots \times R_{bn} . \tag{7}$$

Indeed, if there is a base address that does not belong to the Cartesian product of sets of basic values of the requisites:

$$\exists \ a_b \in A_b \wedge a_b = \{r_1, r_1, \ldots r_n\} \longrightarrow r_{bi} \notin R_{bi} . \tag{8}$$

But this assumption contradicts the requirement that R_{bi} contains all the values of this requisite that are contained in the base addresses.

The set of revoked addresses belongs to the following association of Cartesian products of sets of requisites:

$$A_c \subseteq \begin{cases} R_{c1} \times \overline{R_2} \times \ldots \times \overline{R_n} ; \\ \overline{R_1} \times R_{c2} \times \ldots \times \overline{R_n} ; \\ \ldots \\ \overline{R_1} \times \overline{R_2} \times \ldots \times R_{ci} \times \overline{R_n} ; \\ \ldots \\ \overline{R_1} \times \overline{R_2} \times \ldots \times R_{cn} . \end{cases} \tag{9}$$

Thus, the set of the canceled addresses is contained in the union of the sets made up of Cartesian products of the requisite values. Where each work contains a set of revoked values for at least one requisite and normal sets for the remaining requisites.

IOP Conf. Series: Materials Science and Engineering **537** (2019) 052007 doi:10.1088/1757-899X/537/5/052007

The combination of properties (7) and (9) forms a necessary sign of the normality of the set of addresses $\bar{A}$.

3. Normalization of requisite values

The real set of values of each requisite is wider than the set of its normal values [6]. Therefore, we consider several steps of expanding the set of normal values of the requisite in the direction of the set of real values of the requisite. At each step, the rule (function) of normalization of the extended value of the requisite will be considered, i. e. converting it to one of the normal values.

3.1. Step 0 - syntactic normalization

The values of the requisite can be in the form of a name or the form of a number:

- Requirements for the syntax of requisite values in the form of a name.
- The name must not begin and end with the symbol "space".
- The individual words of the name should be separated by only one space character.
- Between the symbol "hyphen" and the component parts of the words in the name there should not be a symbol "space".
- Between the opening parenthesis and the following word in the name there should not be a space character.
- There must be no space between the closing bracket and the word preceding it in the name.
- There must not be a space between the word next to it.
- Requirements for the syntax of requisite in the form of numbers.
- The number must not begin and end with the space character.
- The number starts with a digit.
- The number in the number can end in one letter, but there should not be a single “space” between the number and the letter.
- The letter that ends the number must not be enclosed in single or double quotes.
- The number can contain no more than 2 numbers, which are separated from each other by the symbol "/".
- There should be no space between the “/” character and the numbers surrounding it.
- The 0-extension of the R_{ex0} requisite value is the set of the requisite value, which is reduced to a normal value by correcting one or more of the name defects listed in the value requirements in the form of a name or number:

$$R_{ex0} \cap \overline{R} = \emptyset. \tag{10}$$

The set of 0-extended values of the requisite is called the union of the normal values of the requisite and the 0-extension of the values of this requisite:

$$\overline{R_{ex0}} \equiv R_{ex0} \cup \overline{R}. \tag{11}$$

0-normalization or 0-normalization of function ν_0 is called function from $\overline{R_{ex0}}$ in $\overline{R}$ or $\overline{R_{ex0}} \xrightarrow{\nu_0} \overline{R}$ is such that:

$$\nu_0(r) \equiv \begin{cases} r, \forall r \in \overline{R} \\ \forall r \in R_{ex0} \exists r' (r' \in \overline{R}) \nu_0(r) = r'. \end{cases} \tag{12}$$

In other words, the function ν_0 leaves the normal values of the requisite unchanged, and each value of the 0-extension sets in correspondence the normal value of this requisite.

3.2. Step 1 – orthographic normalization

Requirements for orthography requisite values in the form of a name.

The spelling of all the words of a name or number must exactly match their spelling to one of the 0 extended requisite values.

1-extension of the values of the requisite R_{ex1} is a set of values of the requisite, each word in which differs from the 0-extended value of the requisite of the value of the requisite by no more than 1 according to the Damerau-Levenshtein metric. At the same time, the values of 1-extension do not belong $\overline{R_{ex0}}$:

$$R_{ex1} \cap \overline{R_{ex0}} = \varnothing. \tag{13}$$

By a set of 1-extended requisites values, we will call the union of 0-extended values and 1-extensions of requisites values:

$$\overline{R_{ex1}} \equiv R_{ex1} \cup \overline{R_{ex0}}. \tag{14}$$

1-normalization or 1-normalization function ν_1 is called function from $\overline{R_{ex1}}$ в $\overline{R_{ex0}}$ or $\overline{R_{ex1}} \overset{\nu_1}{\rightarrow} \overline{R_{ex0}}$ such that:

$$\nu_1(r) \equiv \begin{cases} r, \ \forall r \in \overline{R_{ex0}} \\ \forall r \in R_{ex1} \exists r' \left(r' \in \overline{R_{ex0}}\right) \nu_1(r) = r'. \end{cases} \tag{15}$$

In other words, the function ν_1 leaves 0-extended requisite values unchanged, and each value of the 1-extension assigns a 0-extended requisite value to each.

3.3. Step 2 – normalization of the sequence

Requirements for following words in the value of requisite.

The word order in the meaning of the requisite must exactly correspond to the order of one of the 1-extended requisite values.

2-extension of requisite values R_{ex2} is a set of requisites values that are obtained by permutations of words in the values of a set of 1-extended requisite values, but not belonging to R_{ex1}:

$$R_{ex2} \cap \overline{R_{ex1}} = \varnothing. \tag{16}$$

The set of 2-extended requisite values will be called the union of the set of 1-extended requisite values and 2-extensions of values:

$$\overline{R_{ex2}} \equiv R_{ex2} \cup \overline{R_{ex1}}. \tag{17}$$

2-normalization or 2-normalization function of ν_2 is called function from $\overline{R_{ex2}}$ в $\overline{R_{ex1}}$ or $\overline{R_{ex2}} \overset{\nu_2}{\rightarrow} \overline{R_{ex2}}$ such that:

$$\nu_2(r) \equiv \begin{cases} r \ \ \forall r \in \overline{R_{ex1}} \\ \forall r \in R_{ex2} \exists r' \left(r' \in \overline{R_{ex1}}\right) \nu_2(r) = r'. \end{cases} \tag{18}$$

In other words, the function ν_2 leaves unchanged 1-extended values of the requisite, and each value of the 2-extension is assigned a 1-extended value, i.e. sets the word order of the value of the 2-extension requisite in accordance with the word order of the corresponding value of the set of 1-extended values of this requisite.

3.4. Step 3 – semantic normalization

Requirements for the unambiguity of the semantic value of the requisite.

The value of the requisite *r* must be unambiguous, i.e. there must be no other values of the requisite, defining the same territory, and, therefore, a set of objects:

$$FT(r) = t \wedge \forall r' \left(r' \in R\right) FT\left(r'\right) = t \longrightarrow r' = r. \tag{19}$$

3-extension requisite values R_{ex3} – is the set of values not included in R_{ex2}, but defining the same territories as the values R_{ex2}:

$$R_{ex3} \cap \overline{R_{ex2}} = \emptyset. \tag{20}$$

The set of 3-extended requisite values will be called the union of the set of 2-extended requisite values and 3-extensions of values:

$$\overline{R_{ex3}} \equiv R_{ex3} \cup \overline{R_{ex2}}\,. \tag{21}$$

Let $r(r \in \overline{R_{ex2}})$, then $r'\left(r' \in R_{ex3}\right)FT\left(r'\right) = FT(r)$ is called a synonym for r. And r in turn will be called the original value of the requisite. 3-normalization or 3-normalization function ν_3 is called function from $\overline{R_{ex3}}$ в $\overline{R_{ex2}}$ or $\overline{R_{ex3}} \xrightarrow{\nu_3} \overline{R_{ex2}}$ such that:

$$\nu_3(r) \equiv \begin{cases} r, \ \forall r \in \overline{R_{ex2}} \\ \exists\, r'\left(r' \in \overline{R_{ex3}}\right) \nu_3(r) = r'. \end{cases} \tag{22}$$

In other words, the ν_3 function leaves 2-extended properties of the properties unchanged, and each value of the 3-extension is associated with a 2-extended value, i.e. each synonym is assigned the initial value of the requisite.

4. Address normalization

The real set of addresses is wider than the set of normal addresses. First of all, due to the possibility of expanding the values of requisites. In addition, the extension of the set of addresses beyond the normal set may occur, regardless of whether the components of its requisite values are normalized or not.

The first 4 steps of expanding the set of normal addresses in the direction of the set of real addresses are the steps of expanding the sets of values of the requisites from which the addresses are composed. As in the case of requisites, at each step the rule (function) of normalization of the extended address will be considered, i.e. converting it to one of the normal addresses.

The fifth additional step of expanding the set of normal addresses is characteristic only for the complete address, as the totality of all the values that make up its requisites.

4.1. Step i-normalization based on requisites.

The steps of normalization of addresses with numbers from 0 to 3 are steps based on the normalization of the values of the requisites that make up the address. For brevity, instead of describing each step, we give a general description of these steps, replacing the step number with the symbol *i*.

I-extension of addresses A_i is a set of addresses, the values of at least one of its component requisites are subjected to *i*-extension.

The set of *i*-extended addresses is the union of the set of *i*-1-extended addresses and *i*-extension of addresses:

$$\overline{A_i} \equiv A_i \cup \overline{A_{i-1}}. \tag{23}$$

I-normalization or function of *i*-normalization of ν_i is called function from $\overline{A_i}$ в $\overline{A_{i-1}}$ or $\overline{A_i} \xrightarrow{\nu_i} \overline{A_{i-1}}$, performing *i*-normalization of the values of each requisite that makes up the address that:

$$\nu_i(a) \equiv \begin{cases} a, \ \forall a \in \overline{A_i} \\ \exists\, a'\left(a' \in \overline{A_{i-1}}\right) \nu_i(a) = a'. \end{cases} \tag{24}$$

In other words, the function ν_i leaves unchanged the addresses from the set of *i*-1-extended addresses, and assigns *i*-1-extended address to each *i*- extension address.

4.2. Step 4-semantic normalization of addresses

Requirements for the uniqueness of the meaning of the address.

Address a must be unique, i. e. there must be no addresses defining the same property of the properties.

4-address extension A_4 is a set of addresses not included in $\overline{A_3}$, but defining the same real estate objects as the values $\overline{A_3}$:

$$A_4 \cap \overline{A_3} = \emptyset. \tag{25}$$

The set of 4-extended address values will be called the union of the set of 3-extended address values and 4-extension values:

$$\overline{A_4} \equiv A_4 \cup \overline{A_3}. \tag{26}$$

Let $a(a \in \overline{A_3})$, then $a'(a' \in R_4)F(a') = F(a)$ – is called a synonym for a. A in turn will be called the source address.

4-normalization or function of 4-normalization of ν_4 is called function from $\overline{\mathrm{A}_4}$ в $\overline{\mathrm{A}_3}$ or $\overline{A_4} \xrightarrow{\nu_4} \overline{A_3}$ such that:

$$\nu_4(a) \equiv \begin{cases} a, \ \forall a \in \overline{A_3} \\ \exists\, a'(a' \in \overline{A_3})\nu_4(a) = a'. \end{cases} \tag{27}$$

In other words, the function ν_i leaves the 3-extended addresses unchanged, and each value of the 4-extension assigns a 3-extended address, i.e. each address synonym is assigned an initial address.

Sequential application to synonym of a' all normalization functions results in a normal address belonging to the set of normal addresses $\bar{A}$:

$$\forall a'(a' \in \overline{A_4})\, \nu_0\Big(\nu_1\Big(\nu_2\Big(\nu_3\big(\nu_4(a')\big)\Big)\Big)\Big) \in \bar{A}. \tag{28}$$

5. Conclusion

One of the main tasks of identifying the location of real estate, determining the place of residence of citizens, etc. is the problem of addressing. This problem is complicated by the ambiguity of the concept of address, or rather the values of its requisites.

The relevance of the work is to bring an arbitrary address to normal form.

Based on the set-theoretic approach, a formal definition of the set of normal addresses is formulated.

From the definition of the address as a point in the space of the requisites, the values of the requisites were also normalized. The real set of values of each requisite is wider than the set of its real values. In this regard, the concept of four extensions of the values of the requisite, from which the addresses are composed, is introduced. The fifth additional step of expanding the set of normal addresses is characteristic only for the complete address, as the totality of all the values that make up its requisites.

The proposed method of normalization can be useful not only for the normalization of the concept of address in real estate, but also in other areas.

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] Nagayev RT 2003 *Real Estate: Encyclopedic Dictionary* (Kazan: Publisher GUP «Idel-Press»)
[5] Scheffel O M 2013 *First Full Encyclopedia of Housing and Communal Services* (Moscow: Publisher AST)
[6] Gladkov S L 2015 On the requirements for an intelligent data model *BULLETIN of Moscow University named after Witte S Yu* **2 (10)**