Towards the Theory of Semantic Space

Valerian Ivashenko
Department of Intelligent Information Technologies
Belarusion State University of Informatics and Radioelectronics
Minsk, Republic of Belarus
whitivashenko@bsuir.by

***Abstract*—The paper considers models for investigating the structure, topology and metric features of a semantic space using unified knowledge representation.**

The classes of finite structures corresponding to ontological structures and sets of classical and non-classical kinds are considered, and the enumerability properties of these classes are investigated.

The notion of operational-information space as a model for investigating the interrelation of operational semantics of ontological structures of large and small step is proposed.

Quantitative features and invariants of ontological structures oriented to the solution of knowledge management problems are considered.

***Keywords*—Semantic Space, Neg-entropy, Operationalinformation space, Enumerable sets, Natural numbers, Ackermann coding, Generalized formal language, Enumerable self-founded Hereditarily finite sets, Countable nonidentically-equal Hereditarily finite sets, Multigraph, Hypergraph, Metagraph, Orgraph, Unoriented graph, Quasimetric, Orgraph invariant, Dynamic graph system, Resonator, Graph dimension, Fully-connected orgraph period, Rado graph, Universal model, Stable structure, Operational semantics, Denotational Semantics, Infinite structures, Generalized Kleene closure**

I. Introduction

There are different approaches to the study of topological, metrical and other properties of signs in texts leading to the consideration of corresponding semantic (or meaning (sense)) spaces [56].

Space is convenient because it is connected with some ordinal or metric scale which allows to reduce the cost of solving such cognitive tasks as searching (synthesis) or checking (analysis) the presence of an element (including for the purpose of eliminating redundancy) in a set organized as a space.

Knowledge integration based on unification is necessary both to eliminate redundancy and to compute semantic metrics. For this purpose, the developed model of unified knowledge representation [1], [5] can be adopted.

II. Approaches to the construction of a meaning space

The history of the development of the concept of "meaning space" and the corresponding models are described in the works [2], [11], [32], [56].

As stated in [56], the main approaches to the construction and research of the organization of meaning space include:

- exterior studying the physical nature [30], [33], [48] of processes including thinking processes [29],
- (quantitative) interior using quantitative and soft models, including probabilistic description of processes [11], [34], [35], [42], based on the practice of using words of language [20], [53], [48], [54]
- (qualitative) interior investigating the structure of represented knowledge and its dynamics [12], using formal semiotic models [51].

In some cases, it is possible to combine these elements of these approaches.

The following models and methods are used to construct and investigate the semantic space:

- mathematical models of spaces [37]–[41], [43],
- formal and generalized formal languages [45], [56],
- methods of probability theory [11], [36], [54], [55], [57]
- methods of formal concepts analysis [58], [59], [61]
- other models [3], [4], [45], [46], [49], [51], [54], [56].

Further in the paper we consider the main classes of structures, their attributes and corresponding types of subspaces of the semantic space using unified knowledge representation [5], [12].

III. Unified representation and classification of fully representable finite knowledge structures

At the level of syntax, using syntactic links, it is possible to represent only finite knowledge structures in a unified (explicit) way.

Let us consider the principles of unified representation of knowledge [5], [12] with a structure that is one of finite structures of different kinds. Let us compare a certain class of structures to each kind of finite knowledge structures.

Note that finite structures can be divided into two main types: oriented finite structures and unoriented finite structures [21].

The simplest unoriented finite structures are hereditarily finite sets [63]. The structures are hereditarily class of hereditarily finite sets can be expressed as follows:

$$\emptyset^{(+_1^*)} = H_{\aleph_0}$$

where

$$\binom{0}{A} \stackrel{\text{def}}{=} 2^{\emptyset} = \{\emptyset\}$$

$$\binom{1}{A} \stackrel{\text{def}}{=} \bigcup_{x \in A} 2^{\{x\}} / \binom{0}{A}$$

$$\binom{\iota+1}{A} \stackrel{\text{def}}{=} \left(\binom{\iota}{A} \breve{\bigcup} \binom{1}{A}\right) / \binom{\iota}{A}$$

$$A \breve{\bigcup} B \stackrel{\text{def}}{=} \bigcup_{\langle P, Q\rangle \in A \times B} \{P \cup U\}$$

$$2^{(\emptyset + \Sigma_{x \in A}\{x\})} \stackrel{\text{def}}{=} \bigcup_{\iota \in \mathbb{N} \cup \{0\}} \binom{\iota}{A}$$

$$A^{(+^1_k)} \stackrel{\text{def}}{=} \tau_k(\rho_k(\langle A, A\rangle) \cup \sigma_k(\langle A, A\rangle))$$

$$A^{(+^{\iota+1}_k)} \stackrel{\text{def}}{=} \tau_k\left(\left(\rho_k\left(\left\langle A, A^{(+^{\iota}_k)}\right\rangle\right) \cup \sigma_k\left(\left\langle A, A^{(+^{\iota}_k)}\right\rangle\right)\right.\right.$$

$$A^{(+^*_k)} \stackrel{\text{def}}{=} \bigcup_{\iota \in \mathbb{N}/\{0\}} A^{(+^{\iota}_k)}$$

$$\tau_1(A) \stackrel{\text{def}}{=} A$$

$$\rho_1(\langle A, B\rangle) \stackrel{\text{def}}{=} \emptyset$$

$$\sigma_1(\langle A, B\rangle) \stackrel{\text{def}}{=} 2^{\emptyset + \Sigma_{x \in (A \cup B)}\{x\}}$$

According to Ackermann coding [62], all hereditarily finite sets can be a mutually uniquely matched to natural numbers and thus enumerated [27]:

$$f(S) = 0 + \sum_{x \in S} 2^{f(x)}$$

A generalization of the class of hereditarily finite sets is the class of generalized hereditarily finite sets.

$$A^{(+^*_1)}$$

Generalized hereditarily finite sets can be embedded in (classical non-generalized) hereditarily finite sets:

$$\emptyset \sim 2^{\emptyset}$$

$$a_k \sim 2^{\{\emptyset\}_k}$$

$$g(\emptyset) = \{\emptyset\}$$

$$g(a_k) = \{\{\emptyset\}_k, \emptyset\}$$

$$g(X) = \{g(x) \mid x \in X\}$$

or alternatively:

$$\emptyset \sim d(1) = \{\emptyset\}$$

$$a_k \sim d(2 * k + 1)$$

$$d(k) = \bigcup_{i=1}^{\lfloor \log_2 k \rfloor} \left\{ d\left(\left\lfloor \frac{k}{2^i} \right\rfloor\right) mod 2 \right\}$$

$$d(0) = \emptyset$$

$$d(1) = \{\emptyset\}$$

$$d(2) = \{\{\emptyset\}\}$$

$$d(3) = \{\{\emptyset\}, \emptyset\}$$

$$d(4) = \{\{\{\emptyset\}\}\}$$

$$d(5) = \{\{\{\emptyset\}\}, \emptyset\}$$

$$d(6) = \{\{\{\emptyset\}\}, \{\emptyset\}\}$$

$$d(7) = \{\{\{\emptyset\}\}, \{\emptyset\}, \emptyset\}$$

$$d(8) = \{\{\{\emptyset\}, \emptyset\}\}$$

$$d(9) = \{\{\{\emptyset\}, \emptyset\}, \emptyset\}$$

$$d(10) = \{\{\{\emptyset\}, \emptyset\}, \{\emptyset\}\}$$

$$d(11) = \{\{\{\emptyset\}, \emptyset\}, \{\emptyset\}, \emptyset\}$$

$$d(12) = \{\{\{\emptyset\}, \emptyset\}, \{\{\emptyset\}\}\}$$

$$d(13) = \{\{\{\emptyset\}, \emptyset\}, \{\{\emptyset\}\}, \emptyset\}$$

$$d(14) = \{\{\{\emptyset\}, \emptyset\}, \{\{\emptyset\}\}, \{\emptyset\}\}$$

$$d(15) = \{\{\{\emptyset\}, \emptyset\}, \{\{\emptyset\}\}, \{\emptyset\}, \emptyset\}$$

$$d(16) = \{\{\{\{\emptyset\}\}\}\}$$

$$\ldots$$

$$g(\emptyset) = \{\emptyset\}$$

$$g(a_k) = d(2 * k + 1)$$

$$g(X) = \{g(x) \mid x \in X\}$$

In this way we obtain an ordering of generalized hereditarily finite sets (as example) in accordance with the Ackermann numbering and embedding in hereditarily finite (unoriented) sets.

As for oriented structures (oriented, "ordered" sets), if we take the von Neumann-Bernays-Gödel axiomatics [63] as a basis then with some "traditional" approach (representation of oriented pairs according to K. Kuratowski) [24] an empty string [10], [13], [14], an empty an oriented set [22] cannot be represented as unfounded sets in a theory with the von Neumann-Bernays-Gödel axiomatics [26], [63].

Accepted:

$$x = \langle x \rangle$$

in this case, the oriented pair of K. Kuratovsky:

$$\langle x, y\rangle = \{\{x\}, \{x, y\}\}$$

also

$$\langle x_1, x_2, x_3\rangle \ = \langle\langle x_1, x_2\rangle, x_3\rangle$$

$$\langle x_1, x_2, x_3, x_4\rangle \ = \langle\langle x_1, x_2, x_3\rangle, x_4\rangle$$

$$\langle x_1, x_2, x_3, x_4, x_5\rangle \ = \langle\langle x_1, x_2, x_3, x_4\rangle, x_5\rangle$$

$$\langle x_1, x_2, x_3, x_4, x_5, x_6\rangle = \langle\langle x_1, x_2, x_3, x_4, x_5\rangle, x_6\rangle$$

$$\langle x_1, x_2, ..., x_i, ..., x_{n-1}, x_n\rangle = \langle\langle x_1, x_2, ..., x_i, ..., x_{n-1}\rangle, x_n\rangle$$

$$A^n = \{\langle x_1, x_2, ..., x_i, ..., x_{n-1}, x_n\rangle \mid x_i \in A\}$$

The consequence of this is that strings are conditionally dimensional, that is, the length of a string is not its function, and therefore cannot be calculated uniquely from a string; an empty string cannot be represented by a set in the von Neumann-Bernays-Gödel theory:

$$\langle x, x\rangle = \{\{x\}\} = \langle\{x\}\rangle$$

$$2 = length(\langle x, x\rangle) \neq length(a) = 1$$

$$n = length(\langle x_1, x_2, ..., x_i, ..., x_{n-1}, x_n\rangle) \neq$$
$$length(\langle x_1, x_2, ..., x_i, ..., x_{n-1}, x_n\rangle) = 2$$

During understanding the string length function, if we move from a function (as in the formulas above) to a higher-order function with respect to the set of elements of an oriented set this does not solve the problem:

$$2 = length(\{\langle x, x\rangle, \{x\}\})(\langle x, x\rangle) \neq$$

$$length(\{\langle x, x\rangle, \{x\}\})(\langle\{x\}\rangle) = 1$$

Another consequence of this is that the Cartesian power can exhibit the following non-obvious and non-intuitive properties:

$$\exists A(A = A^1 \supset A^2 \supset A^3 \supset ... \supset A^i \supset ...)$$

Inability to represent the empty string ε when representing strings as oriented sets

Let be:

$$E = \{\varepsilon\}$$

Required:

$$E^n = E$$

We have:

$$E^2 = \{\langle\varepsilon, \varepsilon\rangle\} = \{\{\{\varepsilon\}\}\}$$

$$\varepsilon = \{\{\varepsilon\}\}$$

$$E^3 = \{\langle\varepsilon, \varepsilon, \varepsilon\rangle\} = \{\langle\langle\varepsilon, \varepsilon\rangle\}, \{\langle\varepsilon, \varepsilon\rangle, \varepsilon\}\}\}$$
$$E^3 = \{\{\{\{\{\varepsilon\}\}\}, \{\{\{\varepsilon\}\}, \varepsilon\}\}\}$$
$$\varepsilon = \{\{\{\{\varepsilon\}\}\}, \{\{\{\varepsilon\}\}, \varepsilon\}\}$$

The latter violates the axiom of regularity (foundation), otherwise:

$$E^2 \neq E$$

$$E^3 \neq E$$

The use of non-founded sets is evidence of a transition to non-classical mathematical models

There are approaches to representing strings in von Neumann-Bernays-Gödel set theory by equivalence classes of groupoids (which is complex) over oriented sets or functions (requires the construction of a set of ordinal numbers). In the first case, the representation grows exponentially, and in the second case, it is necessary to use oriented pairs [23], [25] (the number of characters for a string of length n grows no faster than $1 + 14 * n + p(n)$, where $p(n) = 1 + n * (n+3)/2$ – number of characters to represent ordinal numbers). These approaches do not require a transition to non-classical mathematical models. However, a string of one element is not this element

Let's consider another approach to representing strings and oriented sets, which does not require, overcomes the identified difficulties within the framework of classical mathematical models and uses pairs not according to K. Kuratowski, which cannot counter-intuitively have cardinality (length) equal to on

Let us define the concept of disposing of a set

$$1^S \overset{\text{def}}{=} S$$

$$(\iota + 1)^S \overset{\text{def}}{=} \{\iota^T \mid T \subseteq S\}$$

Example.

$$2^{\{x\}} = \left\{1^{\{x\}}, 1^{\emptyset}\right\} = \{\{\chi\}, \emptyset\}$$

$$3^{\{x\}} = \left\{2^{\{x\}}, 2^{\emptyset}\right\} = \left\{\{\chi\}, \emptyset, \left\{1^{\emptyset}\right\}\right\} = \{\{\{\chi\}, \emptyset\}, \{\emptyset\}$$

$$4^{\{x\}} = \left\{3^{\{x\}}, 3^{\emptyset}\right\} = \{\{\{\{\{\chi\}, \emptyset\}, \{\emptyset\}\}, \{\{\emptyset\}\}\}$$

$$5^{\{x\}} = \left\{4^{\{x\}}, 4^{\emptyset}\right\} = \{\{\{\{\{\{\chi\}, \emptyset\}, \{\emptyset\}\}, \{\{\emptyset\}\}\}, \{\{\{\emptyset\}\}\}\}$$

$$4^{\emptyset} = \left\{3^{\emptyset}\right\} = \left\{\left\{2^{\emptyset}\right\}\right\} = \left\{\left\{\left\{1^{\emptyset}\right\}\right\}\right\} = \{\{\{\{\emptyset\}\}\}\}$$

Also, let us define the concept of an individual set.

$$\{x\}_1 \overset{\text{def}}{=} \{x\}$$

$$\{x\}_{\iota+1} \overset{\text{def}}{=} \{\{x\}_\iota\}$$

note that:

$$(\iota + 1)^{\emptyset} = \{\emptyset\}$$

Finite oriented set:

$$\bigcup_{i=1}^{k} \left\{(k - i + 1)^{a_i}\right\}_i$$

The number of characters to represent it is no more than $1 + n * (5 * n + 1)/2 + q(n)$ where $q(n) = 2 * n + 1$ is the number of characters per representation of individual sets of the empty set

Examples:

$$\langle\rangle = \emptyset$$
$$\langle\chi\rangle = \{\{\chi\}\}$$