Construction of Semantic Metric for Measuring the Distance between Ontology Concepts

Viktor Hryhorovych

Lviv Polytechnic National University, S. Bandera street, 12, Lviv, 79013, Ukraine

Abstract
The problem of constructing metrics is crucial for solving the problem of quantitative evaluation of systems of objects of arbitrary nature as a whole, as well as the relations that describe the relationships between the components of these systems.
To assess the relationship between concepts, a metric for non-taxonomic ontology (for an arbitrary semantic network of concepts) is constructed.
The analysis and reasoning of the offered metric is carried out.

Keywords 1
Ontology, semantic network, metrics, size, distance, concept

1. Introduction

Modern information systems simulate subject areas that contain objects and systems of complex structure. The network model is the most adequate for describing the world around it: it reflects objects and systems of objects of arbitrary nature that interact with each other. In fact, any system can be described using a network model.

To quantify systems, the priority is to build an appropriate metric that will define the concept of "system size" and "distance between system elements". This will allow you to accurately describe the systems and the relationships between the elements of the system, to move from their qualitative to quantitative characteristics.

An example of such network systems are ontologies – semantic networks that describe the concepts of certain subject areas.

In the study of systems, when different systems are studied and analyzed, metrics are needed that describe the quantitative characteristics of the systems themselves. For ontologies, these metrics will allow you to compare different ontologies.

When examining the individual elements of a system and the relationships between those elements, metrics that describe the relationships between the individual components of the systems will be important. For ontologies, these metrics will make it possible to estimate the distance between individual concepts of a particular ontology.

Thus, the problem of constructing metrics for quantifying network and hierarchical systems, in particular, ontologies, is important. These metrics will solve a number of problems related to semantic analysis of texts – automatic abstracting of a given text, automatic evaluation of responses to open test tasks, automatic construction of a semantic network for a given text, etc.

This paper considers metrics for network structures described by means of oriented graphs, suitable as semantic metrics for measuring the distance between concepts in an arbitrary (not only hierarchical) ontology.

COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems, April 22–23, 2021, Kharkiv, Ukraine
EMAIL: viktor.grigorovich@gmail.com; viktor.h.hryhorovych@lpnu.ua (V. Hryhorovych)
ORCID: 0000-0002-5828-067X (V. Hryhorovych)

2. Analysis of Literary and Other Sources

Many publications are devoted to metrics for network and hierarchical structures, as well as metrics for evaluating ontologies and concepts in ontologies. The metrics described in the sources can be divided into the following types: morphological metrics and metric for a weighted graph, probability (Bayesian) metrics, metrics for adaptive ontologies, metrics for ontology evaluation and metrics for evaluating relationships between concepts in ontologies.

2.1. Morphological Metrics for Hierarchical Trees

In [1] authors described a set of simple morphological metrics for hierarchical trees based on graph characteristics.

- *Size* = n, where n is the number of vertices.
- The interaction density $R = e / n$ (the ratio of the number of edges to the number of vertices). For the tree $e = n–1$.
- The branching factor for the output *Fan_out*(i) is the number of child vertices of the i-th vertex.

The primary characteristics of the graph are the number of vertices n and the number of edges e.

For a tree to them two more global characteristics – height (depth) and width are added.

- *Height* (*depth*) – the number of levels (the number of vertices in the longest path from the root node to the leaf);
- *Width* – the maximum number of vertices placed on any one level of the tree. The width of the level – the number of tree nodes at this level, while the width of the tree – the maximum width at all levels.

2.2. *q*-Metric for a Weighted Graph and Natural Metric for a Simple Graph

In [2] the metric for the usual graph is given.

Let $L[q] = (X, U; q)$ be an ordinary graph with a weight function q that puts a real number $q(u) > 0$ s a length in accordance with each edge $u \in U$. If Q is a walk, then the sum of $q(Q) \equiv \sum_{u \in Q} q(u)$ for all its edges is called its *q-length*, and simply *length* is the number of edges of the walk (in both cases each edge should be counted as many times as it occurs in the walk).

Number

$$\rho(x, y) \equiv \rho_L^q(x, y) \equiv \min\{q(Q) \mid Q \in \mathbf{Q}(x, y)\}$$

is called the *q-distance* between the vertices $x, y \in X$ of the weighted graph $L[q]$. Here $\mathbf{Q}(x, y)$ is the set of all simple trails from x to y. If $x = y$, then Q is a trail of zero length and its *q-length* $q(Q) \equiv 0$, and if the vertices x and y are separated, then $\rho(x, y) \equiv +\infty$.

The *q-distance* satisfies the three axioms of the Fréchet metric.

$$\forall\, x, y \in X\, [\rho(x, y) = 0 \Leftrightarrow x = y],$$

$$\forall\, x, y \in X\, [\rho(x, y) = \rho(y, x)],$$

$$\forall\, x, y \in X\, [\rho(x, y) + \rho(y, z) \geq \rho(x, z)],$$

that is the metric on the set X. In the partial case, when all $q(u) = 1$, i.e. when the *q-distance* of each trail is equal to its normal length, the metric $\rho(x, y) \equiv \rho_L^q(x, y)$ of the graph $L[q]$ is called the *natural metric* of the simple graph $L = (X, U)$.

2.3. Probability (Bayesian) Metrics

In the work [3] a probabilistic approach based on Bayes' theorem is proposed for estimating hierarchical structures in the construction of expert systems.

The metric for such systems is based on Bayes' theorem: the probability of some hypothesis H in the presence of certain evidence E that confirms this hypothesis (i.e., when events occurred E), is calculated based on the a priori probability of this hypothesis without evidence-confirmation E and the probability of evidence for conditions that the hypothesis is true or false.

$$P(H \mid E) = \frac{P(HE)}{P(E)} \quad \Rightarrow \quad P(HE) = P(H \mid E) \cdot P(E) = P(E \mid H) \cdot P(H)$$

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

where

$P(H)$ – a priori probability of the hypothesis H;

$P(H \mid E)$ – the probability of the hypothesis H when an events occurred E (a posteriori probability);

$P(E \mid H)$ – the probability of occurrence of events E with the truth of hypothesis H;

$P(E)$ – probability of occurrence of events E.

In [4] author develops the approach proposed in [3].

Let $G = (V, E)$ be a connected graph, and u and v be its two different vertices. Then the distance between the vertices u and v will be the length of the shortest walk, which is denoted by $\rho(u, v)$. In this case, all axioms of the metric are fulfilled.

It is known that a binary relation, which can be given by an adjacency matrix, mutually uniquely represents each graph. The elements of the adjacency matrix $A(G)$ have the form

$$a_{ij} = \begin{cases} 1, & \text{if the vertices with numbers } i \text{ and } j \text{ are compatible} \\ 0, & \text{otherwise} \end{cases}$$

The rank of a graph G is called the rank of its adjacency matrix, denoted by $rank(G)$.

If u is some vertex of the graph $G = (V, E)$, the value of $e(u) = \max \rho(u, v)$, $v \in V$ is called the eccentricity of the vertex u. The diameter of the graph is called the maximum eccentricity among its vertices and denote $d(G) = \max e(u)$, $u \in V$.

In this case, we could use the diameter of the graph as a measure and build a metric of graphs based on their diameters – which is equivalent to one of the morphological metrics.

As an example of hierarchy, author in [4] cites the tree of human diseases and their relationships depending on the causes of their occurrence and mechanisms of their development. The basis of the metric for such systems uses the Bayes' theorem. The use of this approach in building an expert system for the medical knowledge base MYCIN is given in [3]; a similar approach to the analysis of crystal structures of chemical compounds is described in [5].

It should be noted that the MYCIN system operates with the concept of "degree of certainty". The procedural rules in it are formulated in the form of "IF … THEN … WITH CERTAINTY P", where the degree of certainty – "about the same as we call the conditional probability $P(H \mid E)$ – the probability of hypothesis H, provided that the event E occurred" [3]. When building the MYCIN system, medical experts proposed rules and indicated the degree of confidence in each rule in the range from 1 to 10 – such expert assessments and became a stellar certainty for the relevant procedural rules. Thus, the set of procedural rules of such a system can be described by means of an oriented graph, each edge of which has weight – the conditional probability of transition $E \rightarrow H$, i.e. the probability of hypothesis H under the condition that event E occurred (Figure 1).

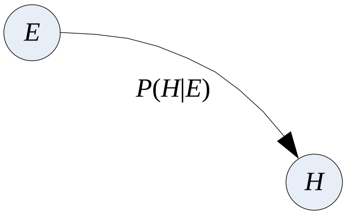

Figure 1: Illustration to the probability metric

Obviously, this approach allows us to estimate only some parts of a given oriented graph and is not suitable for comparing different graphs, because the total probability of a complete system should be equal to one: $P(\, G = (V, E)\,) = 1$.

2.4. Metrics for Evaluating Ontologies

Many approaches to defining metrics for evaluating ontologies have been described in the literature. Approaches based on the linguistic characteristics of the essences of ontologies, or approaches based on the internal structure of the ontology (hierarchy of classes and relationships between classes) are mainly used.

Paper [6] proposes to use machine learning to define semantically similar concepts of different ontologies. In [7], a metric is introduced to evaluate the division of ontologies into separate modules; dividing the ontology into separate modules will improve their use in semantic web applications. The paper [8] describes the methods of alignment of ontologies for their further comparison, using metrics based on the distance between letter lines. [9] is a review article that analyzes various metrics for ontologies. In [10], metrics were introduced and an algorithm for normalizing ontologies was described, which is based on the structure of ontology classes. Metrics for ordering ontologies based on keywords are proposed in [11] and [12]. [13] introduces connection metrics based on ontology references to external classes. [14] describes ontology metrics based on the number of root classes, the number of leaf classes, and the average depth of the class hierarchy. In [15], a semantic metric was proposed based on the number of semantic sections of the knowledge base, the number of uncoordinated (incompatible) subsets in the knowledge base and the ratio of the number of uncoordinated sets to the number of elements in the knowledge base. In [16-17], a methodology based on certain meta-properties of entities is described, which allows to identify inconsistent relations in the ontology. The paper [18] describes primitive metrics and metrics of complexity of ontologies; primitive metrics contain morphological parameters of the ontology: the total number of classes, relations, paths; complexity metrics are the average number of relationships per concept and the average number of paths per concept. In [19] an overview of ontology evaluation methods is given. In [20] it is proposed to interpret ontology as a two-level system that contains lexical and conceptual levels. In [21] the systematics from Wikipedia is estimated based on maps of relations of tokens and concepts. In [22] the taxonomy constructed based on comparison with the reference taxonomies constructed from Wikipedia is estimated. In [23] it is proposed to present an ontology in the form of a vector space and on this basis to calculate the similarity of ontologies at the lexical level and the level of relations. A similar approach is proposed in [24], which describes the approach to estimating the constructed taxonomy based on comparison with the reference; metrics are used: content quality – based on the number of label overlaps between two taxonomies and structural – based on the number of hierarchical relationships between the respective elements of different taxonomies.

2.5. Metric for Evaluating of Connections between Concepts in Adaptive Ontologies

Metrics based on adaptive ontologies for semantic (for example, classification problems) and sign (for example, search for relevant precedents) problems are proposed in [25-28].

For semantic problems, the distance between precedent and situation is defined as the sum of the distances between the "most important" concepts of precedent and the current case. The most important concept corresponds to the center of gravity of the conceptual graph, through which the adaptive ontology is presented. A maximum of three "most important" concepts of the conceptual graph are considered. In this case, we obtain three weight centers of the i-th precedent and three weight centers of the current situation s^1, s^2, s^3. The distance between the precedent and the current situation is defined as

$$d(pr, s) = \arg\min \sum_{n=1}^{3} d_n ,\ d_n = d(pr_i^j, s^k),\ j = 1,2,3,\ k = 1,2,3$$

– from nine different distances $d\left(pr_i^j, s^k\right)$, j=1,2,3; k=1,2,3 such three are chosen that their sum would be minimal. The amount received will be the distance between the precedent and the current situation.

This metric does not evaluate all paths from one concept to another.

2.6. Ontology to the Description of Metrics

The paper [29] describes the development of an ontology to organize information about Metrics and its potential application to identify and manage the project metrics CTSA (Clinical and Translational Science Award). The goal is to maintain an integrated database of all metrics used by CTSA components. The ontology is presented as a conceptual schema of entity-relationship data.

3. System Analysis and Justification of the Problem

Ontology is an "explicit specification of conceptualization" [30]. Ontology – a formal and explicit definition of conceptualization, which contains

- concepts (concept of subject area),
- their definition,
- hierarchical organization of concepts,
- the relationship between concepts,
- axioms for formalizing definitions and relationships.

Here, conceptualization means the introduction of abstract objects for an abstract, simplified description of the world around us.

To automate the semantic analysis of texts, it is necessary to introduce metrics on network structures to evaluate quantitative characteristics that describe

- ontologies in general and
- the relationship between concepts in a particular ontology.

The available metrics analyzed in Section 2 can be divided into two types.

1. Metrics describing the aggregate integral characteristics of the ontology.

They allow you to compare different ontologies.

Integral characteristics are such characteristics of ontology that

- evaluate the ontology as a whole;
- allow you to compare different ontologies with each other.

Such characteristics are, for example, the following characteristics of the ontology graph:

- number of nodes,
- number of edges,
- diameter,
- maximum, minimum, average degree of the node, etc.

2. Metrics describing the characteristics of the relationship between the concepts of one ontology.

They allow you to compare concepts in algorithms for solving problems such as automatic abstracting, searching and evaluating texts, and more.

All relationships between concepts can be divided into two categories.

- Taxonomic relations.

They describe the hierarchical relationships "is-a" ("… is a kind of…") between concepts, ie the relationship "subset - set", "set - superset".

- Non-taxonomic relationships.

They describe non-hierarchical relationships, such as the *used-in* relationship ("term_1 occurs in term_2") or the *uses-of* relationship ("term_1 uses term_2", or "term_1 refers to term_2").

Available metrics of this kind estimate the distance between concepts in assuming the existence of a single path from one concept to another.

If the ontology is a taxonomy, that is, all connections between concepts are hierarchical, then this assumption is valid and the specified metrics are true.

However, if there are non-taxonomic relationships between concepts (for example, "term_1 refers to term_2"), then such a metric will not be true because they assume the existence of many paths from one concept to another.

Metrics based on many relationships between concepts should take into account

- distance as the number of transitions in the oriented graph of the ontology on the way from one node-concept to another,
- the number of such paths.

Thus, the task is to build a metric that will allow us to estimate the distances between concepts in the case of non-taxonomic relationships in the ontology.

4. Presenting the Main Material

Let us build a metric for network structures, suitable for measuring the distance between concepts in the case of non-taxonomic ontology.

This metric will allow us to estimate the distance between the concepts of ontology in the assumption of the existence of many paths from one concept to another. Such network structures are described by arbitrary oriented graphs.

Everywhere in the paper we will consider the terms "node" and "vertex" are equivalent.

4.1. Metric for Concepts of Non-Taxonomic Ontologies

Definition of Inverse-Additive Metric

Denote by N_i – the number of transitions from concept A to concept B on the i-th path, i=1, …, K , where K is the number of different paths that can be passed on the oriented graph of some ontology from concept A to concept B.

Define the distance $R(A, B)$ between the concepts A and B as follows

$$\frac{1}{R(A,B)} = \sum_{i=1}^{K} \frac{1}{N_i}$$

Example

Assume that there are two paths between concepts A and B. The first path contains one transition, and the second – two (Figure 2).

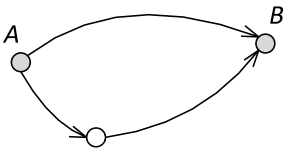

Figure 2: Illustration to the example of inverse-additive metric

Then the distance $R(A, B)$ between them

$$\frac{1}{R(A,B)} = \frac{1}{1} + \frac{1}{2} = \frac{3}{2}, \quad R(A,B) = \frac{2}{3}$$

Analogy

An analogy to this metric is the rule for calculating the electrical resistance for a series and parallel connection. Based on Ohm's law, for series connection of resistors R_1 and R_2 the total resistance will be

$$R = R_1 + R_2,$$

For parallel connection

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$

4.2. Reasoning

As is known, the metric is based on the notion of distance. The distance $d(x, y)$ is an unambiguous, non-negative, real function d: X × X → $\mathbb{R}$, defined for $\forall$ $x, y \in$ X, which satisfies three axioms of the metric.

1) $d(x, y) = 0 \Leftrightarrow x \equiv y$ (Axiom of Identity)
2) $d(x, y) = d(y, x)$ (Axiom of Symmetry)
3) $d(x, z) \le d(x, y) + d(y, z)$ (Axiom of a Triangle)

We prove that these axioms hold for the specified metric.

Axiom of Identity

In this case, $R(A, A) = N$, where N is the number of transitions from node A to node A, N=0.

The first axiom is true. □

Axiom of Symmetry

It should be noted that in the general case for an oriented graph there could be no symmetry in the interpretation of the second axiom of the metric. That is, the rule $R(A,B) = R(B,A)$ cannot be fulfilled, because the number of transitions from node A to node B will coincide with the number of transitions from node B to node A only if there is a cycle $A \to B \to A$ and the distance (number of transitions) on the path $A \to B$ is equal to the distance on the path $B \to A$.

The introduction of pairs of symmetric relations, for example, for ontology – a pair of connections *uses-of* – *used-in*, allows fulfilling the axiom of symmetry for the proposed metric in the following interpretation

$$R_{used-in}(A,B) = R_{uses-of}(B,A)$$

If we consider a pair of mutually symmetric relations (the *uses-of* relation is symmetrical to the *used-in* relation), then the second axiom is true. □

Axiom of a Triangle

Consider an oriented graph for which we substantiate the axiom of a triangle (Figure 3).

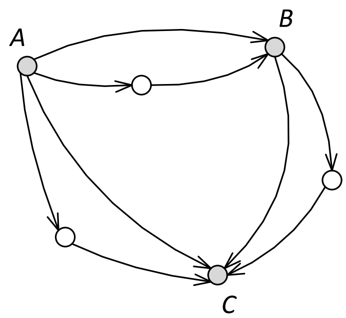

Figure 3: Illustration to the axiom of a triangle for inverse-additive metric

Let us mark

$R(A,B) = R(\widehat{AB})$ is the distance between nodes A and B. $\widehat{AB}$ – all paths from node A to node B.

$R(B,C) = R(\widehat{BC})$ is the distance between nodes B and C. $\widehat{BC}$ – all paths from node B to node C.

$R(A,C)$ is the distance between nodes A and C.

$R(\widehat{AC})$ is the distance between nodes A and C along the path $\widehat{AC}$ that does not pass through node B.

$R(\widehat{ABC})$ is the distance between nodes A and C along the path $\widehat{ABC}$ that passes through node B.

According to the definition of this metric, after the path $\widehat{AC}$ is added to the paths $\widehat{AB}$ and $\widehat{BC}$, the distance $R(A,C)$ will be determined by the formula

$$\frac{1}{R(A,C)} = \frac{1}{R(\widehat{AC})} + \frac{1}{R(\widehat{ABC})}$$

Then

$$R(A,C) = \left(\frac{1}{R(\widehat{AC})} + \frac{1}{R(\widehat{ABC})}\right)^{-1} = \left(\frac{1}{R(\widehat{AC})} + \frac{1}{R(\widehat{AB}) + R(\widehat{BC})}\right)^{-1} =$$

$$= \left(\sum_{i=1}^{K_{\widehat{AC}}} \frac{1}{N_i} + \left(\left(\sum_{i=1}^{K_{\widehat{AB}}} \frac{1}{N_i}\right)^{-1} + \left(\sum_{i=1}^{K_{\widehat{BC}}} \frac{1}{N_i}\right)^{-1}\right)^{-1}\right)^{-1}$$

Since

$$R(A,B) = R(\widehat{AB}) = \left(\sum_{i=1}^{K_{AB}} \frac{1}{N_i}\right)^{-1} = \left(\sum_{i=1}^{K_{\widehat{AB}}} \frac{1}{N_i}\right)^{-1}$$

and

$$R(B,C) = R(\widehat{BC}) = \left(\sum_{i=1}^{K_{BC}} \frac{1}{N_i}\right)^{-1} = \left(\sum_{i=1}^{K_{\widehat{BC}}} \frac{1}{N_i}\right)^{-1}$$

then

$$\frac{1}{R(A,C)} \geq \frac{1}{R(A,B) + R(B,C)}$$

Hence

$$R(A,C) \leq R(A,B) + R(B,C)$$

That should have been proved.□

An Example for the Axiom of a Triangle

Consider the case shown in Figure 3. Here

$$R(A,B) + R(B,C) = \frac{4}{3}$$

$$\frac{1}{R(A,C)} = \sum_{i=1}^{6} \frac{1}{N_i} = \frac{1}{1} + \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3} + \frac{1}{4} = \frac{24+8+3}{12}$$

$$R(A,C) = \frac{12}{35} \leq R(A,B) + R(B,C) = \frac{4}{3}$$

4.3. Finding the Distance between Concepts

Fragment .owl-file that contains the ontology terms from computer science, is shown in Figure 4.

```
<!-- urn:absolute:Ontology_02#Algorithm -->

<owl:NamedIndividual rdf:about="urn:absolute:Ontology_02#Algorithm">
    <rdf:type rdf:resource="urn:absolute:Ontology_02#Term"/>
    <UsedIn rdf:resource="urn:absolute:Ontology_02#Program"/>
    <definition rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        a finite sequence of commands, the execution of which leads to the result
    </definition>
    <keyword rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        Algorithm
    </keyword>
</owl:NamedIndividual>

<!-- urn:absolute:Ontology_02#Program -->

<owl:NamedIndividual rdf:about="urn:absolute:Ontology_02#Program">
    <rdf:type rdf:resource="urn:absolute:Ontology_02#Term"/>
    <UsesOf rdf:resource="urn:absolute:Ontology_02#Algorithm"/>
    <definition rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        an algorithm written using a programming language
    </definition>
    <keyword rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        Program
    </keyword>
</owl:NamedIndividual>
```

Figure 4: Fragment .owl-file that contains the ontology terms from computer science

Concept definitions are contained in the `Individuals` section, each concept begins with the tag `owl:NamedIndividual`. The concept ID is contained in the attribute `rdf:about`. The concepts associated with the current concept by *uses-of* relation are listed in the subsections that correspond to the `UsesOf` tags, and their identifiers are the value of the `rdf:resource` attribute. Concepts related to the current concept by *used-in* relation are listed in the subsections that correspond to the `UsedIn` tags, and their identifiers are also the value of the `rdf:resource` attribute.

Thus, parsing an owl file will reveal the connections between the concepts.

With the help of known algorithms for traversing an oriented graph, we can find all the paths from one given concept to another, and based on the found paths – to calculate the distance between these concepts. In fact, the transport flow problem should be solved in an oriented graph of ontology from the source node corresponding to the first concept to the receiving node corresponding to the second concept.

5. Discussion and Analysis of the Obtained Results

There are some problems with the fact that the ontology is actually an oriented graph, the nodes (vertices) of which are concepts, and the edges are connections between concepts.

In the general case, more than one path connects the two vertices of an oriented graph of ontology.

5.1. The Problem of Symmetry of Connections between Concepts – the Problem of Symmetry of Connections for an Oriented Graph

The problem is that, in general, for an oriented graph of ontology the axiom of symmetry is not fulfilled

$$R(A, B) \neq R(B, A)$$

Here, the distance between the concepts is determined based on morphological metrics – as the number of transitions *N* between the nodes of the oriented graph of the ontology on the way from concept *A* to concept *B*

$$R(A, B) = N$$

This problem can be solved by constructing pairs of mutual inverse relationships, for example, for each relationship "uses" *uses-of* build an inverse relationship "used" *used-in*.

There are two approaches to constructing non-taxonomic relationships between concepts in ontologies.

The **first approach** provides the maximum completeness of the description of the subject area in the ontology. According to this approach, we represent as fully as possible the connections between the concepts of the subject area through the relationships between the concepts in the ontology.

If we consider non-taxonomic relations, then each pair of concepts represented by neighboring nodes in the semantic network of the ontology will be represented by a pair of *used-in* and *uses-of* relations.

At the same time, the large number of connections between concepts makes the ontology too "cumbersome", which complicates its practical use in application systems.

For example, let us build a semantic network for the concepts PROGRAM and ALGORITHM based on the following definition: "a program is a record of an algorithm using a programming language".

Definition of the term PROGRAM uses (*uses-of*) the term ALGORITHM, and the term ALGORITHM is used in (*used-in*) the definition of the term PROGRAM (Figure 5).

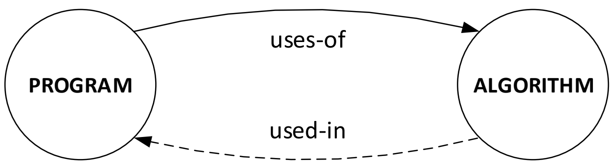

Figure 5: Illustration to a semantic network with "uses-of" and "used-in" connections between concepts

The distance between concepts *A* and *B* in such a semantic network can be defined as the distance that takes into account the existence of two paths that correspond to the symmetrical relationships "uses" (*uses-of*) and "used" (*used-in*) between these concepts

$$\frac{1}{R(A,B)} = \frac{1}{R_{used-in}(A,B)} + \frac{1}{R_{uses-of}(A,B)}$$

It is this approach that ensures symmetry of the connections, and

$$R(A,B) = R(B,A)$$

The **second approach** provides simplicity of ontology and, as a consequence, - speed of calculations and efficiency at construction of application systems. Since the *used-in* relations are derived from the *uses-of* relations, it is possible not to define them explicitly in the ontology. At the same time, we deliberately neglect the completeness of the subject area in order to simplify the ontology.

For our example, the semantic network corresponds to the phrase "Definition of the term PROGRAM uses (*uses-of*) the term ALGORITHM" (Figure 6).

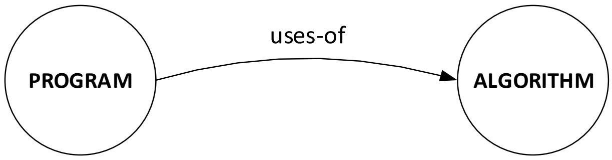

Figure 6: An illustration of a semantic network that has only "uses-of" connections between concepts

For this approach, the axiom of symmetry is not fulfilled

$$R(A,B) \neq R(B,A)$$

5.2. The Problem of the Axiom of a Triangle for an Oriented Graph

As with the axiom of symmetry, there are problems with the axiom of a triangle: in the general case, the axiom of a triangle for an oriented graph is not fulfilled.

Again, consider the case where the distance between the concepts is determined based on the morphological metrics – as the number of transitions N between the nodes of the oriented graph of ontology on the way from concept A to concept B

$$R(A,B) = N$$

Consider the following example (Figure 7).

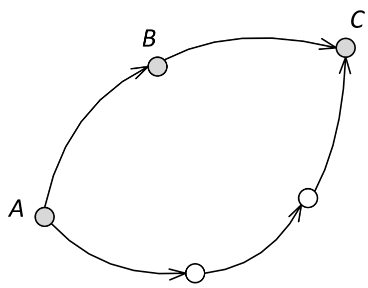

Figure 7: Illustration to the problem of the axiom of a triangle for morphological metric in an oriented graph

For this example

$$R(A, B) = 1,\ \ R(B, C) = 1,\ \ R(A, C) = 3,$$

Obviously, in this case the axiom is a triangle

$$R(A, C) \le R(A, B) + R(B, C)$$

is not fulfilled.

Natural metric [2] allows solving this problem. The distance between nodes is defined as $R(A, B) = \min\{N_i\}$, where N_i is the number of transitions from concept A to concept B on the i-th path, i=1 …, K, where K is the number of different paths in which you can go on the oriented graph of some ontology from concept A to concept B.

Then, given that there are 2 paths from node A to node C: path $\widehat{ABC}$, which passes through the concept B, and the path $\widehat{AC}$, which does not pass through the concept B, and $R(A, B) = 1$, $R(B, C) = 1$, $R(\widehat{ABC}) = 2$, $R(\widehat{AC}) = 3$, $R(A, C) = \min\{R(\widehat{ABC}), R(\widehat{AC})\} = \min\{2, 3\} = 2$, we obtain the triangle rule

$$2 = R(A, C) \leq R(A, B) + R(B, C) = 2.$$

Another solution to the problem may be the introduction of inverse-additive metrics, which also takes into account that there are two paths from concept A to concept C.

Then

$$R(A, B) = 1, \quad R(B, C) = 1$$

$$\frac{1}{R(A,C)} = \frac{1}{R(\widehat{ABC})} + \frac{1}{R(\widehat{AC})} = \frac{1}{R(A,B) + R(B,C)} + \frac{1}{R(\widehat{AC})} = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$$

Hence

$$R(A,C) = \frac{6}{5} = 1{,}2$$

Now the triangle rule is fulfilled

$$1{,}2 = R(A,\ C)\ \leq\ R(A,\ B)\ +\ R(B,\ C) = 2$$

5.3. The Problem of Cycles – the Cyclical Connections between Concepts

The problem that arises due to the presence of recursive cycles.

A graph of the relationships between the following concepts is shown in Figure 8.

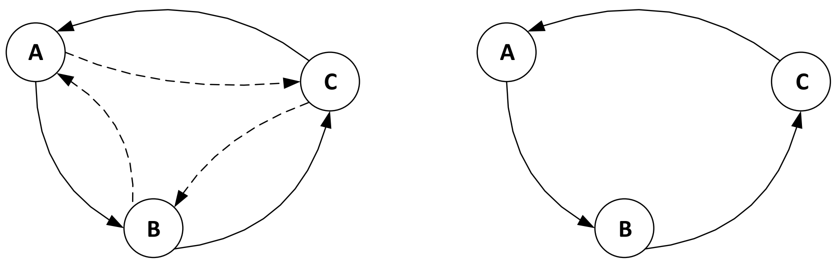

Figure 8: Illustration to semantic networks that have recursive connections between concepts; on the left – "uses-of" and "used-in"; on the right – "uses-of" only

Then the distance $R(A, B)$ between concepts A and B in this semantic network can be defined as follows

$$\frac{1}{R(A,B)} = \frac{1}{R_{AB}(A,B)} + \frac{1}{R_{BCA}(B,A)}$$

where $R_{AB}(A,B)$ is the distance from node A to node B along the path $A \rightarrow B$, $R_{BCA}(B,A)$ is the distance from node B to node A along the path $B \rightarrow C \rightarrow A$.

The distance $R(A, B)$ determined in this way takes into account the presence of two paths that connect nodes A and B.

6. Conclusions

A metric is introduced that allows estimating the distances between concepts in an ontology when there are many paths from one concept to another. The proposed metric allows fractional values of distances between nodes-concepts of the oriented ontology graph and provides a correct estimate for non-taxonomic relations between concepts.

The proposed metric will solve problems related to semantic analysis of texts - automatic abstracting of a given text, automatic evaluation of answers to open test tasks, automatic construction of a semantic network for a given text, etc.

7. References

[1] N. E. Fenton, S. L. Pfleeger, Software Metrics: A Rigorous and Practical Approach, 2nd. ed., International Thompson Computer Press, 1997.

[2] A. A. Zykov, Fundamentals of graph theory, Science, Moscow, 1987 (original title: Зыков А. А. Основы теории графов. Наука, Москва, 1987).

[3] Chris Naylor, Build your own PC expert system, Sigma Press, 1983.

[4] G. S. Tesler, Metrics and norms in the hierarchy of categorical semantics and functions, Mathematical machines and systems (original title: Г. С. Теслер. Метрики и нормы в иерархии категориальных семантик и функций, Математичні машини і системи) 2 (2005) 65-68.

[5] V. Yu. Velychko, Solving analytical problems in discrete media by inference methods by analogy (original title: Величко В. Ю. Розв'язання аналітичних задач в дискретних середовищах методами виведення за аналогією), Ph.D. thesis, Institute of Cybernetics, Kyiv, 2004.

[6] Duy Hoa Ngo, Zohra Bellahsene, Remi Coletta, Generic Approach for Combining Linguistic and Context Profile Metrics in Ontology Matching, in: Proceedings of the ODBASE'2011: 10th International Conference on Ontologies, DataBases and Applications of Semantics, Crete, Greece, 2011, pp.800-807.

[7] Alsayed Algergawy, Samira Babalou, Birgitta Konig-Ries, A New Metric To Evaluate Ontology Modularization, in: Proceedings of the 2nd International Workshop on Summarizing and Presenting Entities and Ontologies, Co-located with the 13th Extended Semantic Web Conference, Greece, 2016. URL: http://ceur-ws.org/Vol-1605/paper4.pdf.

[8] Giorgos Stoilos, Giorgos Stamou, Stefanos Kollias, A String Metric for Ontology Alignment, in: Proceedings of the International Semantic Web Conference ISWC 2005: The Semantic Web – ISWC, 2005, pp. 624-637.

[9] J. García, F, J. García-Peñalvo, R. Therón, A Survey on Ontology Metrics, in: Proceedings of the World Summit on Knowledge Society WSKS: Knowledge Management, Information Systems, E-Learning, and Sustainability Research, 2010, pp. 22-27.

[10] Denny Vrandecic, York Sure, How to Design Better Ontology Metrics, in: The Semantic Web: Research and Applications, Springer-Berlag, 2007, pp. 311-325.

[11] Harith Alani, Christopher Brewster, Nigel Shadbolt, Ranking Ontologies with AKTiveRank, in: Proceedings of the International Semantic Web Conference, ISWC, 5th International Semantic Web Conference (ISWC), 2006. URL: http://www.cbrewster.com/papers/Alani_ISWC06.pdf.

[12] Harith Alani, Christopher Brewster, Metrics for Ranking Ontologies, in: Proceedings of the 4th Int. EON Workshop, 15th Int. World Wide Web Conference, 2006. URL: http://ceur-ws.org/Vol-179/eon2006alanietal.pdf

[13] Anthony Orme, Haining Yao, Letha Etzkorn, Coupling Metrics for Ontology-Based Systems, IEEE Software, 2006, pp. 102-108.

[14] Haining Yao, Anthony Orme, Letha Etzkorn, Cohesion Metrics for Ontology Design and Application, Journal of Computer Science 1(1) (2005) 107-113.

[15] Yinglong Ma, Beihong Jin, Yulin Feng, Semantic oriented ontology cohesion metrics for ontology-based systems, The Journal of Systems and Software, Elsevier, 2009.

[16] Nicola Guarino, Chris Welty, An Overview of OntoClean, in: The Handbook on Ontologies, Berlin, Springer-Verlag, 2004, pp. 151-172.

[17] Nicola Guarino, Chris Welty, Evaluating Ontological Decisions with OntoClean, in: Communications of the ACM, ACM Press (2002) 61-65.

[18] Zhe YANG, Dalu Zhang, Chuan YE, Evaluation Metrics for Ontology Complexity and Evolution Analysis, in: Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE'06), 2006.

[19] Joe Raad, Christophe Cruz, A Survey on Ontology Evaluation Methods, in: Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, At Lisbon, Portugal, 2015.

[20] A. Maedche, S. Staab, Measuring similarity between ontologies, in: Knowledge engineering and knowledge management: Ontologies and the semantic web, Springer, Berlin Heidelberg (2002) 251-263.
[21] S. P. Ponzetto, M. Strube, Deriving a large scale taxonomy from Wikipedia, AAAI 7 (2007) 1440-1445.
[22] P. Treeratpituk, M. Khabsa, C. L. Giles, Graph-based Approach to Automatic Taxonomy Generation (GraBTax), 2013, arXiv preprint arXiv:1307.1718.
[23] Elias Zavitsanos, George Paliouras, George A. Vouros, Gold standard evaluation of ontology learning methods through ontology transformation and alignment, IEEE Trans. on Knowl. and Data Eng. 23 (2011) 1635-1648.
[24] V. Kashyap, C. Ramakrishnan, C. Thomas, A. Sheth, TaxaMiner: an experimentation framework for automated taxonomy bootstrapping, International Journal of Web and Grid Services 1(2) (2005) 240-266.
[25] V. Lytvyn, Knowledge base of intelligent decision support systems (original title: В. В. Литвин. Бази знань інтелектуальних систем підтримки прийняття рішень), Lviv Polytechnic Publishing House, Lviv, 2011.
[26] D. Dosyn, V. Lytvyn, Yu. Nikolsky, V. Pasichnyk, Intelligent systems based on ontologies (original title: Д. Г. Досин, В. В. Литвин, Ю. В. Нікольський, В. В. Пасічник. Інтелектуальні системи, базовані на онтологіях), Civilization, Lviv, 2009.
[27] V. Lytvyn, A method of entering metrics to determine the distance between text documents (original title: Литвин В. В. Спосіб введення метрики для визначення відстані між текстовими документами), Information systems and networks 621 (2008) 162-171.
[28] V. Lytvyn, V. Vysotska, D. Dosyn, O. Lozynska, O. Oborska, Methods of building intelligent decision support systems based on adaptive ontology, in: Proceedings of the IEEE Second International Conference on Data Stream Mining & Processing, 2018, pp. 145-150.
[29] Dagobert Soergel, Olivia Helfer, A Metrics Ontology. An intellectual infrastructure for defining, managing, and applying metrics, Knowl Organ Sustain World Chall. Perspect. Cult. Sci. Technol. Shar. Connect. Soc. (2016) 333-341.
[30] Gruber T. R. A translation approach to portable ontologies, Knowledge Acquisition 5(2) (1993) 199-220.