DOI: https://doi.org/10.15407/kvt205.03.005

UDC 681.3.

GRITSENKO V.I.[1], Corresponding Member of NAS of Ukraine,
Director of International Research and Training Center
for Information Technologies and Systems
of the National Academy of Sciences of Ukraine
and Ministry of Education and Science of Ukraine
ORCID: 0000-0003-4813-6153, e-mail: vig@irtc.org.ua
GLADUN A.Ya.[1], PhD (Engineering),
Senior Researcher of the Department of Complex Research
of Information Technologies and Systems
ORCID: 0000-0002-4133-8169, e-mail: glanat@yahoo.com
KHALA K.O.[1], Researcher of the Department of Complex Research
of Information Technologies and Systems
ORCID: 0000-0002-9477-970X, e-mail: cecerongreat@ukr.net
MARTÍNEZ-BÉJAR R.[2], PhD (Informatics),
Professor at the Department of Information and Communication Engineering
and Artificial Intelligence
ORCID: 0000-0002-9677-7396, e-mail: rodrigo@um.es
[1] International Research and Training Center
for Information Technologies and Systems of the National Academy of Sciences
of Ukraine and Ministry of Education and Science of Ukraine
40, Acad. Glushkov av., Kyiv, 03187, Ukraine
[2] University of Murcia
CP 30180 Bullas, Spain

SEMANTICAL SIMILARITY EVALUATION METHOD OF CONCEPTS FOR COMPARISON OF ONTOLOGIES IN APPLIED PROBLEMS OF ARTIFICIAL INTELLIGENCE

Introduction. *The expediency of reapplication of ontology in applied intelligent information systems (IIS), which are focused on functioning in the open Web environment on the basis of Semantic Web technologies, is substantiated in the work. Features of ontology storage and management platforms and their metadata are analyzed. Possibilities of searching in ontology repositories and their reuse in IIS are considered. The mechanisms of ontology search based on semantic processing of their metadata, analysis of ontology structure using metrics of semantic similarity between their concepts related to the current user task are presented.*

The purpose *of the paper is to develop algorithms and methods for evaluating semantic models, which consist in combining qualitative (ontological) representation of knowledge with quantitative (numerical) evaluation of ontologies and their parameters (semantic proximity, semantic distance, semantic affinity) aimed at finding similarities different ontologies.*

Methods. *Methods of ontological analysis of objects of the subject area, theoretical and multiple approaches to determine the degree of closeness of two objects by comparing their properties (feature matching) and traditional methods of statistical analysis are used to solve the tasks set in the work.*

Results. *The proposed method of estimating semantic similarity allows on the basis of semantic analysis of natural annotations of metadata both ontologies and data (including Big Data) to perform the task of their interpretation and selection to the problem to be solved by the applied IIS or application. The obtained results allow to create original IIS for artificial intelligence in economics, medicine, national security, defense and social sphere.*

Conclusion. *The proposed original approach to the evaluation and analysis of metadata (ontologies, data) is based on semantic analysis of metadata and determining the semantic similarity of structural data models (ontologies, data) and the formation of a ranked set of related ontologies to solve problems of artificial intelligence. The application of methods for defining semantically similar concepts is presented as a tool for semantic comparison of the structure of ontologies, which were found in the repository under formal conditions, with a poorly structured NL-description. At present, there is no generally accepted standard for presenting metadata, so the proposed methods of analysis of N annotations are the most adequate means of comparing the semantics of ontologies, data with the problems for which they can be used.*

Keywords: *semantic similarity, formal ontology model, metadata, metadata standards, intelligent information system, ontology repository.*

INTRODUCTION

The construction and implementation of modern intelligent information systems (IIS) based on the formalization and reuse of knowledge is a promising area of research and practical application of artificial intelligence methods. The development of intelligent information technologies (IIT) provides for the creation of a new class of IIS based on a formalized representation of knowledge about the subject area (SA). These new information technologies must be able to analyze the environment based on its figurative perception, by using models of knowledge about its objects, phenomena, processes; obtaining the necessary data to achieve the set tasks; structuring this data into certain categories, allow computer processing of these models for solving applied task.

We suggest to solve the problem of reliability, relevance, persistence of information resources (IR) based on a semantically approach to the analysis of metadata that accompanies the information resources and analytically information processing about these resources. Metadata contains a large amount of information about IR, including significant descriptive textual information, the understanding of which by machines would improve the problem of relevance of the applied information objects (data, ontologies, texts etc.).

Nowaday, a huge number of ontologies have already been created in various SA. These ontologies often use one of the standardized presentation languages (OWL or RDF) that designed for multiple repeated reuse, but they have different complexity, structure and quality [1]. Modern means allow searching for the desired ontology among them only according to some formal parameters (for example, by keywords) and not at the level of their semantics. Therefore, it is quite often easier to create a new ontology than searching and selecting a readymade one, which is extremely inefficient and time-consuming. The solution to such problems could be provided by ontology repository [2], which processes knowledge not of the ontologies themselves, but knowledge of ontologies.

In our work, the interest in creating ontology repositories is related the need to search for ontologies that could be reused to create artificial intelligence applications. Until now, such a search was carried out by users manually, but the ensuring automated generation of estimates of available ontologies, the availability of a single metadata standard for describing ontologies and their processing will greatly facilitate this work and increase the persistence of search and selection. The first step in solving the problem will be the semantic binding of the ontology to a certain SA (or several SA), to assess its depth and structural complexity.

Ontologies and dictionaries are key resources for creating interoperable metadata in the Semantic Web. To simplify and accelerate the task of identifying and use relevant ontologies, we using the idea of ontology repositories, they are formed as ontological systems, which has been currently implemented in many international projects [3]. Ontological systems operate wiht ontology models and form of their repositories .

Today, there is an urgent need to use specialized ontologies repositories of different classes, each of which can be focused on different types of *user needs specifications (user profile)*, *different ontology profiles* selected according to a certain topic and *different requirements of organizations*, as they can not be submitted as a common and unique implementation.

Each ontology repository is a separate information system with its own user interfaces and APIs. Ontologies use dynamically-variable languages such as OWL, OWL2, SKOS (Simple Knowledge Organization System), RDF Schema etc. SKOS provides an easy, intuitively comprehensible standard language for developing and disseminating the new knowledge management systems and transferring them to the Semantic Web. This language can be used separately or in combination with a formal knowledge representation language such as OWL.

Semantic technologies based on logic, databases and the Semantic Web can solve the problem of efficient access to data and integration of data that have been created both today and long ago — for decades and centuries. The international project Open Ontology Repository [4] is an initiative to develop and deploy a new interaction infrastructure, called an open ontology repository (OOR).

In this regard, global search, update and inference in repositories are today a difficult and generally poorly implemented task. As a result, it becomes quite difficult effectivelly search and reuse of existing ontologies. Thus, there is a need for knowledge engineers in the ontology analysis tool to be able to evaluate a particular ontology for reuse.

Using a certain taxonomy, the user iteratively identifies the SA of his interests, and the search is not reduced to rearrangement of keywords, although they are used at the initial stage. If more than one ontology is found for user purposes, their parameters must be evaluated. The values of these parameters are calculated automatically (or automated) for each ontology when it is placed in the repository. In particular, the parameters may be the completeness of the ontology, the number of classes and instances in it, the date of creation, the authors, the confirmation certificate (authenticity) of the knowledge contained in it. The user needs to specify the relative weight of the various criteria.

The lack of mechanisms and standards for storing and presenting ontologies affects the process of recognizing, identifying and accessing ontological resources. Thus, the urgent problem is to create new methods to support efficient access and reuse of ontologies with greater scalability and more reliable infrastructure the so-called *ontology repositories*.

Ontology repositories require additional knowledge of ontologies as metadata, which must also be managed together with the ontologies in the repository. Metadata of ontologies — knowledge that contains information about the possibilities of working with the ontology, a description of the ontology itself, ways of its functioning, structure, methods of knowledge extraction, interaction of components etc.

The purpose of the paper is to develop algorithms and methods for evaluating semantic models, which consist in combining qualitative (ontological) representation of knowledge with quantitative (numerical) evaluation of ontologies and their parameters (semantic similarity, semantic proximity, semantic distance, semantic affinity) and aimed at finding different similarities.

RESEARCH OBJECTIVE

Modern intellectual applications require the use of external sources of knowledge, which determines the relevance of the problem of searching ontologies. For reusing ontologies from repositories, it is necessary to develop tools for semantizing their search and analysis, which provide a comparison of metadata and ontology structure with the current user's tasks that require knowledge from these ontologies. For this purpose it is proposed to use such standards for presenting of metadata on ontologies, that allow structuring this information, and methods for determining the semantic proximity between concepts as a tool for quantifying the similarity between ontologies and natural description of the user's task.

BASIC CONCEPTS AND CHARACTERISTICS OF DATA REPOSITORIES AND ONTOLOGIES

In the historical tour, repositories specializing in the preservation of ontologies were developed on the basis of the concept of data warehouses. There are many different values and definitions of data repositories in the literature, so first we will discuss what we will mean by the data warehouse in the future.

The **data repository** is a set of the digital data that is available to one or more entities (or users of systems) for various purposes (training, administrative procedures, research) and has the characteristics offered in [5]:

• the content is placed in the repository by its creator or owner – a third party;

• the repository architecture allows you to manage both content and metadata;

• the repository offers a minimum set of basic services, such as receiving, searching, access control;

• the repository must be stable and reliable, well maintained and well managed.

The term "Data warehouses" became popular in the early 1990s. The purpose of the data warehouse is to analyze the stored data for management decision-making. Data is periodically entered into this data repository, and is usually only added to existing ones. The data repository, however, does not necessarily have to support data warehouse functionality such as analysis.

Like the data repository, there are also many different definitions for the term "knowledge base"(KB). However, in the general case, the knowledge base is a centralized repository of knowledge artifacts. Typically, the KB can use ontologies to formally submit the content and classification schemes, but it can also include unstructured or informal information presented in natural language or procedural code. In addition, unlike the data repository, usually the purpose of the KB is the possibility of automatically deductive inference from the accumulated knowledge.

The Semantic Web community is interested in using repositories to preserve semantic content (for example, ontologies).

Initial projects to organize a base of existing ontologies proposed the creation of library systems that proposed various functions for the managing, adapting and standardizing of ontology groups. These systems are important tools for grouping and reorganizing ontologies for further reuse, integration, maintenance, display and versioning. They defined a model for evaluating the library system based on functionality. Examples of library ontology systems are: WebOnto, Ontolingua, DAML Ontology Library System, SchemaWeb etc. Today, efforts are being made to create ***ontology repositories***. The ontology repository is most similar to the library ontology system defined by [6], but there are some differences.

The term "ontology repository" can be considered as a development of the term, which came from the classical understanding of data repositories [7]. Otherwise, you can rely on the following understanding into the ontology repository and their corresponding control systems.

The ontology repository (OR) is a set of ontologies accompanied by metadata describing individual ontologies and sets of ontologies, their properties and the relationships between them.

Metadata can characterize various aspects of ontologies related to access to ontologies and related to their preservation. The general requirement is that the ontology repository should support the entire ontology lifecycle, from the ontology development process to its use in any intelligent application through specialized tools and tasks. In addition, one of the most important tasks of the ontology repository is the long-term preservation of knowledge.

Ontological KB are a key element of IIS based on Semantic Web. The increase in the number of such applications determines the rapid increase in the number of ontologies that are suitable for use in more than one IIS. In this regard, the problem of organizing effective ontological repositories knowledge bases — ontologies repositories is relevant.

For automatically processing shared knowledge, the consortium *W3C* has developed common standards for their presentation: *RDF (Resource Description Framework) and OWL (Ontology Web Language).*

The basic construction of the *RDF* language is a statement given by the triple <subject> <predicate> <object>. Using *URI* to specify subjects and properties allows to be bind individual statements into complex semantic networks that have a single interpretation in an open environment.

The most common form of saving ontologies is an OWL file. When reading such a file in RAM, a model (set of statements) is created, with which further work is performed. However, this approach has the disadvantages associated with information processing: a significant increase in RAM costs when working with large ontologies due to the full load of the OWL file, as well as a significant increase in loading time of OWL files as the number of ontologies used increases. This does not allow the use of this approach when creating large IIS. An alternative to it is construction of RDF-repositories based on relational databases, which are also designed to store ontological information, but in a different view.

The RDF-repository is an information system designed to store RDF-triplets and execute queries to them. The main functions of the repository are to manage the functions of saving and searching for ontologies in a relational database, provision of a software interface for retrieving knowledge from ontologies stored using the structured query language SPARQL or a special API, and support for the administration of preserved ontologies: adding, deleting, modifying and allocating access rights [8].

Efficient storage must meet the following requirements:

• high productivity — minimization of query execution time;

• minimum memory consumption (disk space) for saving ontologies;

• universality of the approach — the possibility to preserve ontologies of any structure.

There are two basic approaches to the organization of saving ontologies in RDF-repositories:

1) using a single table to store all triplets;

2) mapping the hierarchy of ontological entities (classes, properties, instances) into the RDB scheme.

A feature of another approach is the definition of the DB scheme in accordance with the specific SA, which allows you to optimize the execution of queries. The implementation of this approach for large ontologies involves the creation of a large number of DB tables with complex relationships between them. We present the generalized scheme of RDF-storage as follows (Fig. 1).

The use of RDF repositories is also directly related to the use of metadatafor searching and reusing the ontological knowledge contained in such repositories.

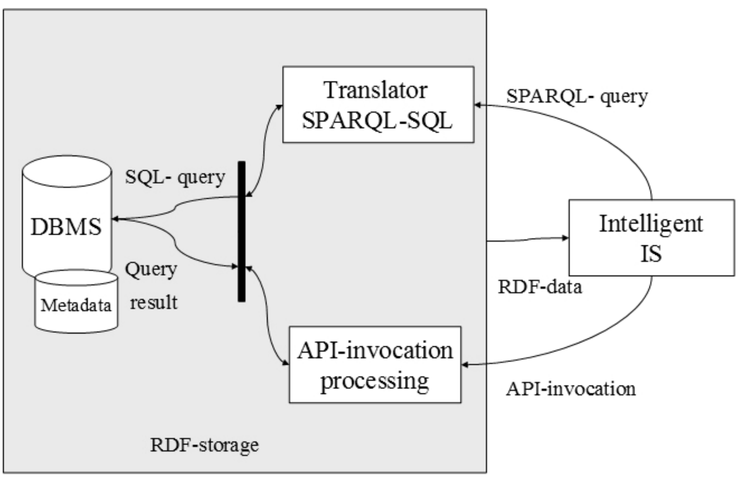

Fig. 1. Generalized *RDF-storage schema*

Metadata is information that makes data useful and should provide access to information [9]. The metadata is defined as structured data that containing the characteristics of the entities they describe for the purposes of their identification, retrieval, evaluation and management of them. Metadata can be used to determine the semantics of information, and therefore to improve its searching and sampling, understanding and using. For example, considers the use of ontologies and thesauruses for semantic annotation of IR and their elements, which is the basis for machine learning and knowledge acquisition from data [10]. Depending on the purposes of annotation, ontologies of different complexity can be applied (from controlled dictionaries and glossaries to ontologies with complex relationship of inversion, non-intersection etc.).

For today in Ukraine, three international standards concerning metadata (ISO 15489-1: 2016 [11], ISO 15836-1: 2017 [12], ISO 15836-2: 2019 [13]) are accepted as national standards by confirmation method [14], [15]. These standards can be applied to display the main properties of Big Data with provision of a common universal language for creating and analyzing metadata, as well as describing the general properties of metadata elements required for basic interoperability between different programming languages and their SA.

For metadata semantic analysis in ontology repositories, we use NL annotations, which are part of the metadata. Semantic processing of metadata information allows to obtain implicit knowledge about the data itself.

ONTOLOGIES EVALUATION CRITERIA

Many different and alternative criteria can be used to evaluate ontologies. Analysis of the literature on ontological analysis [16, 17] allowed to form a set of the most common criteria for evaluating Web-ontologies in the ontologies repository and principles for creating qualitative ontologies.

Gómez-Pérez [18] introduces two terms of ontology verification and validation to describe the ontology evaluation: ontology verification deals with the creation of a correct ontology, that is, ensures that its definitions implement the correct requirements. Ontology validation refers to the content (values) of the definitions of how real they model the SA for which the ontology was created. Ontology validation is an important part of assessing the quality of an ontology and is usually a way to guarantee the correctness of the knowledge encoded in the ontology.

But most approaches to ontology validation require close collaboration with SA experts and cannot be performed automatically [19].

Other criteria for ontologies evaluating are:

• ***Sequence:*** fixation of both the logical sequence (i.e. no contradictions can be logically deduced) and the sequence between the formal and informal description (i.e. comments and formal correspondence of descriptions);

• ***Completeness:*** all knowledge that is expected to be in the ontology, either explicitly declared or deduced from the ontology;

• ***Brevity:*** when the ontology is free from any unnecessary, useless, or excessive axiom;

• ***Extensibility:*** the ability to add new definitions without changing the already established semantics;

• ***Sensitivity:*** refers to how small changes in the axiom change the semantics of the ontology.

Thomas Gruber identified the following criteria:

• ***Clarity:*** the ontology should effectively provide the meaning of the term being defined. The Definitions must be objective. Sometimes the definition can be established using logical axioms. Wherever possible, the definition should take precedence over the description. All logical objects must be documented in natural language.

• ***Consistency:*** the statements must be correct. At least, certain axioms must be logically consistent. In addition, natural language documentation should be consistent with formal statements.

• ***Extensibility:*** the ontology should cover the conceptual foundation for the range of expected tasks and its presentation should be processed so that everyone, if necessary, could expand and specialize the ontology. At the same time, new terms can be introduced without the need to revise existing axioms.

Other researchers [20] identified the following criteria:

• ***coverage of a specific domain*** and completeness, complexity and level of detail through this coverage;

• ***clarity*** for people (users);

• ***legality and reasonableness***;

• for performing an ontology evaluation, the following should be developed: specific uses, scenarios, requirements, applications and ontology data sources;

• sequence;

• completeness;

• the type of logical conclusions for which they can be used;

• adaptability and reuse for broader purposes;

• display at the top level or other ontologies.

Gangemi [21] defines the following criteria:

• ***Cognitive ergonomics:*** these are the fundamental properties of ontology, thanks to which it can be easily understood and managed;

• ***Transparency:*** properties of the ontology, thanks to which it can be analyzed in detail, with a rich formalization of conceptual sets and motivations;

• ***Computational integrity and efficiency:*** ontology properties, thanks to which it can be successfully/easily processed by a reasoner (reasoner, logical inference mechanism, classifier etc.);

• ***Integrity level mark:*** ontology properties through which it maintains a certain order of criteria that are accepted as qualitative characteristics.

• ***Flexibility:*** ontology properties that allow it to be easily adapted to many applications and evaluations.

• ***Consent to the examination:*** properties of the ontology, due to which it is suitable for use by one or more users;

• ***Consent to the procedures of extension, integration, adaptation etc.:*** ontology properties, thanks to which it can be easily understood and managed for multiple use and adaptation;

• ***Universal availability:*** ontology properties that allow you to easily access it for effective use by the application;

• ***Organizational suitability:*** ontology properties that allow it to be easily deployed within the organization and it has the correct annotation of the context.

For different applications, the importance of these criteria may vary significantly, but it is advisable to provide the values of these criteria in the metadata of the repository ontologies (or at least allow them to be automatically generated formed based on other metadata).

For qualitative evaluation and comparison of ontologies in the repository, it is advisable to use quantitative evaluations of its individual aspects with their subsequent integration. Such aspects of the ontology (applicable in the assessment) are:

• ***Dictionary.*** An ontology dictionary is a set of all names in that ontology, URI, or character constants that denote a data type or identify a language. This aspect deals with the different sets related to the URIs or literals used.

• ***Syntax.*** Web ontologies can be described using many different syntaxes: RDF/XML, OWL etc.

• ***Structure.*** Web ontology is described by an RDF or OWL graph. The ontology structure is a graph. The structure can dramatically change even the semantic description of the same ontology. These differences can be assessed based on this aspect.

• ***Semantics.*** A consistent ontology describes a non-empty, usually infinite set of possible models. The ontology semantics is a general characteristic of all these models. This aspect of semantics determines the distinctive hallmarks (features) of the ontology.

• ***Submission.*** This aspect captures the relationship between structure and semantics. Aspects of representation are typically evaluated by comparing the metrics calculated on a simple RDF or OWL graph with the features of possible models as defined by the ontology.

• ***Context.*** This aspect is about the features of an ontology when compared to other artifacts that may be present, such as an application using ontology, a data source describing ontology, different data representations within an ontology, or formalized ontology requirements in the form of competence questions.

ONTOLOGY REPOSITORIES CONCEPTUAL STRUCTURE

From a technical point of view, the practical implementations of ontology repositories differ from each other. However, components and services at the conceptual level are reusable for various technical solutions. As a result, consider the conceptual structure for ontology repositories (Fig. 1). Based on the various implementations of ontology repositories [22], it is possible to determine a set of relevant components and services that should be built into a scalable and reliable structure.

The Ontology Repository Management System (ORMS) is a system for storing, organizing, updating and retrieving knowledge from an ontology repository. An example of using such a program is the Generic Ontology Repository Framework (GORF) [23], which contains a special module to support ontology knowledge. GORF is based on the experience gained in the implementation of the ontology repository "Ontology" and ontology metadata vocabulary (OMV). One of the main requirements for ORMS is scalability and the ability to interact with other repositories — for example, using Web-services.

Table 1. **Sample for OMV metadata element:**

Name	**The name of the metadata element (entity)**
Type	The type of ontological primitive used to represent an element in OWL: Class, ObjectProperty or DatatypeProperty
Identifier	The unique identifier used for this element
Occurrence Constraint	One of the following: required, optional or extensional
Category	Content-dependent category to which belongs the element
Definition	A short definition of the goal, which can be described in detail in the comment tags
Domain	OMV element subject area (for OWL properties)
Range	OMV entity element rank (for OWL properties)
Cardinality	OMV element power (MIN: MAX)
OMV version	The version of OMV in which the element is presented
Comments	Element's detailed description

The ontology repository framework includes such conceptual levels:

1. Access to knowledge, according to the concept of the Semantic Web, for people and machines with support for individual views of users and various visualizations to perform personalized queries.

2. Processes and services for processing the knowledge accumulated in the repository, providing analysis of the quality of ontologies, their comparison, evaluation of their adequacy and reliability.

3. Organization of knowledge processing in the repository taking into account the modular approach [24] for reuse and using metamodel based on open standards: metadata, considered as a metamodel, helps to improve the availability and reuse of ontologies and provide useful information about resources to support maintenance.

4. Ontology repository management should support search, view and navigation in the repository with support for content semantics and the use of various specialized ontologies [25].

Creating metadata for ontologies based on standards and metaontologies. Both specialized standards and universal standards for describing metadata can be used for this purpose. An example of a universal standard is Dublin Core, which is used for various types of documents, the use of which is difficult because it does not take into account the specifics of ontologies. An example of a specialized standard is OMV [26], which is the first metadata standard for ontologies and related entities. It is formalized as an OWL ontology. OMV [27] represents metadata model for ontologies that reflects key aspects of ontology metadata information, such as origin and availability (Table 1). Metadata categories are distinguished between the following three limitations of occurrence for metadata elements, according to their impact on the evaluation of reuse of the described ontological content: 1) mandatory — mandatory metadata elements; any missing entries in this category result in an in-complete description of the ontology; 2) additional — important, but not mandatory, facts of metada-ta; 3) advanced — specialized metadata elements that are not considered core part of the metadata schema.

USING OF SEMANTIC SIMILARITY BETWEEN CONCEPTS FOR ANALYSIS OF METADATA IN ONTOLOGY STORAGE

In order to compare not only the properties of ontologies that related to their quality, validity, scope etc., but also their relevance to a particular user task, it is advisable to the quantitative characteristics of ontology evaluation, namely to compare the semantic component of their metadata with the metadata of the task to be solved or with the metadata of those data whose processing is the user's goal (for example, Big Data metadata). A direct comparison of ontologies will give a more accurate result, but the problem is that:

1) direct comparison of ontologies is a time-consuming and resource-intensive task;

2) in many cases, the purpose of such a comparison is to find the ontology that is most pertinent to the user's task and therefore it is necessary to compare the ontology repositories with a structured or natural (unstructured) description of the task.

Using formal ontology estimates allows you to filter out ontology repositories, among which you search and proceed to the analysis of metadata of these ontologies. Using the same metadata standards to describe tasks, data, and ontologies greatly simplifies the task and compares only semantically related fields.

However, it should be borne in mind that different terms and terminological systems can be used in NL descriptions, and therefore there is a valuation problem the semantic proximity between two independently created NL entities used in metadata.

Semantic similarity and semantic proximity metrics between the concepts of ontology and their parameters. Semantic similarity is a special case of semantic proximity. Semantic similarity takes into account only the hierarchical relationships between the elements of the ontology, while semantic proximity allows us to analyze arbitrary relationships in the ontology. Some researchers suggest that the assessment of similarity in semantic networks should be considered involving only taxonomic connections, excluding other types of connections; but the relationships between the parts can also be seen as attributes that influence the definition of similarity. Many similarity criteria have been identified in the scientific literature, but they are rarely accompanied by an independent characterization of the phenomenon they measure: their value lies in their usefulness for a particular task.

Semantic similarity concepts (SSCs) is a fuzzy set that includes a set of concepts for which the quantitative value of semantic proximity with the selected concept is above a given threshold [28]. Measures for determining the semantic proximity of concepts based on ontologies use a variety of semantic characteristics of these concepts — their properties (attributes and relationships with other concepts), the mutual position in ontological hierarchies. The complexity of the problem of constructing a set of SPC in an ontology is associated with its poor scaling: the increase in the number of concepts in the ontology and the complexity of its structure significantly increase the search space.

The semantic distance between concepts depends on the length of the shortest path between the vertices and the general specificity of the two vertices. The shorter the path from one node to another, the more semantic similar they are. If there are several paths between the elements, the shortest of them is used. The length of the shortest path in this taxonomy between the corresponding concepts, which is determined by the number of vertices (or edges) in the shortest path between the two corresponding vertices of the taxonomy, taking into account the depth of the taxonomic hierarchy (the smaller length of the path between the vertices, the semantically closer in distance). Unfortunately, uniform distance in taxonomy is difficult to determine and even more difficult to control.

The similarity of concepts is also related to their information content. *Information content* of the concept *c* can be quantified as: $-log\ p(c)$, the higher the concepts in the hierarchy, the lower its informativity. Thus, the higher the level of abstraction of the concept, the less its information content. If there is a unique top concept in a taxonomy, then its information content is 0. One of the key factors in the similarity of the two concepts is the degree to which they share the information specified in the IS-A taxonomy by a highly specific concept that applies to both of these concepts. The edge-counting method takes this into account indirectly, because if the minimum path of IS-A connections between two nodes of a graph is long, it means that it is necessary to rise high in taxonomy to more abstract concepts in order to find the smallest upper bound — the concept to which both concepts are analyzed. Such quantitative characterization of information provides a new way of measuring semantic similarity based on the extensional concepts.

In the process of processing natural language information, there is often a need to measure the similarity of words rather than concepts. Using to represent words from the set *W* through the set of concepts in the taxonomy, which are the meanings (contents) of the word *w*, a function $s(w)$ such that $s: W \rightarrow C$, that is $s(w \in W) = \{c_k \in C,\ k = \overline{1,m}\}$, it can determine

$$sim\_w(w_1, w_2) = max\, sim(c_i, c_j) \text{ where } c_i \in s(w_1),\ c_j \in s(w_2).$$

The similarity of words is evaluated by finding the maximum information content over all concepts for which both words can be an instance. This allows you to create sets of *semantically similarity words* (SSW), that is words whose semantic distance between which is less than the selected threshold value.

Many ontology-based proximity measures are based on Tversky's set-theoretic approach [29], which determines the degree of similarity of two objects by *feature matching*. The similarity measures *S(a, b)* between objects *a* and *b* is a function of the three sets of properties of these objects *A* and *B* — their intersection $A \cap B$ and complements *A–B* and *B–A*. The attributive proximity measures is based on the proximity of the values of common attributes of concepts, whose ranges are literals, numbers, rows and other data types.

Using these proximity measures allows to evaluate the similarity of the values of concept parameters (the properties of data instances of ontology classes), which in semanticized Wiki-resources correspond to the values of semantic properties that are not links to other Wiki-pages.

The analysis of the existing approaches to the quantitative assessment of the semantic similarity of concepts shows the appropriateness of using taxonomies for this purpose and the distance in these taxonomies between the concepts whose similarity is assessed of their common superclass. Taking into account other types of ontological relationships between concepts and comparing their semantic properties allow to refine these estimates in accordance with the specifics of SA. Existing approaches and metrics for assessing semantic similarity, as well as methods of their application to unstructured natural language texts are considered in more detail in [30].

Measures of semantic similarityreflect the semantics of ideas about SA, which is reflected in a certain ontology. Thus, it can be assumed that for ontologies that reflect a similar view of SA, the sets of SSC should also be similar. The following approach to searching for ontologies in repositories based on the analysis of their metadata and their structure is based on this assumption.

METHOD OF SEMANTIC SIMILARITY EVALUATION OF ONTOLOGY TO THE USER'S TASK

For searching a pertinent ontology, it is necessary to solve the problem inverse to the search task using semantic similarity metrics: a set of SSCs is formed for a set of ontologies that have passed the first stage of filtering, and we compare these sets with the NL description of the user task. The criterion of similarity is the number of matches (the number of points by coincidence) of the studied ontology with the description of the current task, i.e. the most pertinent will be the ontology whose SSC contains the most matches with the description of the task. It should be noted that the final choice of the most pertinent ontology should be made by the user who is offered a ranked set of ontologies with the highest scores.

The developed method of evaluation of semantic similarity concepts (semantic closeness of ontology to user request) is aimed at solving the problem of searching for natural language text in ontologies at user's request. It should be noted that the search for ontology is usually performed at the first stage of IIS development, for which you need to find an ontology, which can already determine the basic requirements for ontology — its scope, representation language and expressive complexity (it is important not to use too complex and large ontologies in tasks that do not require it, because it complicates the calculation and increases the processing time), but it is difficult to assess the relative importance of different concepts of SA.

Proposed method of assessing the semantic similarity of ontologies comprises the following main stages.

Stage I. It is to determine the conformity of the basic concepts of the ontology under investigation with the query conditions. At this stage, the set K is formed — a set of keywords that characterize the most important concepts of the SA task. The ontology repository searches for these keywords. Ontologies that do not contain all relevant concepts are not considered further. If no ontology in the repository meets the conditions of the query, you need to use another repository or make changes to the set of K (for example, specify a concept in another language or delete some concepts). If more than one ontology is found, you must proceed to stage II.

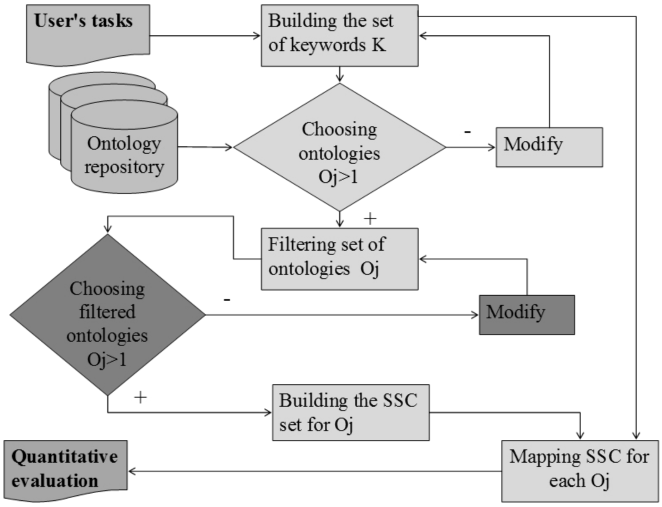

Fig. 2. Algorithm for assessing semantic similarity of the ontology to the problem of artificial intelligence to be solved by the user

Stage II. It is filtering the set of ontologies according to main characteristics. Filtering of a set of ontologies is carried out by volume, presentation language, expressiveness etc. For example, a user can search for taxonomies submitted in OWL 2 that contain between 300 and 1,000 concepts. If no ontology in the repository meets all the selected query conditions, you need to use another repository (filtering is performed in stages I and II, but as a result has different solutions) or weaken the search conditions according to the formal characteristics of the desired ontology. If at this stage more than one ontology is found, then you need to pass to the III stage.

Stage III. It is an assessment of the semantic similarity of the investigated ontology of the user's task description. For each concept $k_i \in K, i = \overline{1,n}$ required in each of the selected ontologies $o_j \in filter(O) \subseteq O, j = \overline{1,m}$ build a set of SSC:

$$t_{i_j} \subseteq T(o_j) \in K, i = \overline{1,n}, j = \overline{1,m},$$

where t_{i_j} is a concept (class or instance of a class) from a non-empty set *T*, taken from *m* terms for an ontology o_j from a non-empty set of selected ontologies of n elements and compare it with NL description of the user's task. Comparison parameters and thresholds for the concept to belong to the SSC are completely determined by the specifics of the task, but these quantitative evaluation are based on the metrics for assessing semantic similarity, analyzed above. Combinatorial methods, linguistic and statistical analysis tools can be used to compare SSC with the task description [31]. The results of comparing individual SSCs for each ontology are summed and normalized (methods and parameters of calculations also depend on the specifics of the task). An algorithm was developed using this method (Fig. 2).

As a result of using this algorithm, each ontology receives a quantitative evaluation of the proximity to the description of the query:

$$x(o_i) = f(s(t_{i_j}, K)), \quad i = \overline{1,n}, \quad j = \overline{1,m},$$

where s(t_i, K) is the result of a comparison t_i from K.

Visualization of the described algorithm is given at Fig. 2.

MODULE FOR SEMANTIC SIMILARITY EVALUATION OF CONCEPTS IN INTELLECTUAL APPLICATIONS OF CYBERSECURITY

Today, the issue of information security (IS) is becoming a cornerstone in the activities of each organization or individual. Information security means the protection of information and the entire organization from intentional or accidental actions that lead to damage to its owners or users.

Cybersecurity is the process of applying security measures to ensure the confidentiality, integrity and availability of data. Cybersecurity protects resources (information, computers, servers, businesses, individuals). Cybersecurity is designed to protect data during its exchange and storage. Such security measures include access control, training, audit and risk assessment, testing, management and authorization security.

Despite the high interest in big data, their analytics for cybersecurity and the availability of various technological means of their storage and processing them, today there are no relevant methods for selecting a pertinent subset of external big data units based on semantic description of metadata suitable for this task. To solve this problem, a module for assessing the semantic similarity of concepts has been proposed, which will be a part of the Cybertrack, the main task of which is to monitor, search and analyze social media on cybersecurity and Big Data technology, in particular, including Elastic Stack components and graph DBMS Neo4j.

The proposed method is implemented in the structure of the module for assessing the semantic similarity of concepts in intellectual applications used to solve information security problems of organizations based on the recognition, selection and interpretation of Big Data units. The module consists of the following units (Fig. 3).

The unit for forming of an array of researched subjects ontologies, which provides selection of key features (words) that characterize the most important concepts of SA tasks, and the definition of the corresponding features in the researched technologies.

The unit for processing the natural description of the user's task, which provides analysis of the text of the user's task and the formation of a set of keywords (thesaurus) of the task.

The Big Data and Metadata Repository is storing external data unit and metadata for their subsequent semantic analysis.

Besides, the analysis is performed at the request of user with the following characteristics: unstructured or weakly structured natural language text, the presence of explicitly or implicitly described SA, input information, processing methods and desired results to be achieved as a result of such processing. Also, the task description in the query may contain references to similar developments and their shortcomings, which need to be eliminated in a new solution of the task. An example of a weakly structured description of the user's task may be a request for the discovery of scientific work, which indicates the thematic direction of scientific research and technical developments, keywords, justification for the feasibility of the work, purpose and objectives, its relevance, tasks solving, experience and refinement authors, structure and stages of work, expected results etc. — use of standardized unit's names allows for better comparison such a description with the metadata of resources and data.

The ontology set filtering unit provides the first stage of selecting relevant ontologies to solved the user's task.

The unit for determining quantitative evaluation of semantic similarity of the investigated ontology of user's task description (selection of ontology closest to the context with user's request). Combinatorial methods, linguistic and statistical analysis tools can be used to compare SSC with the description of the task. The results of comparing individual SSC for each ontology are summarized and normalized by calculation methods and parameters, which also depend on the specifics of the task. Analysis of the semantic similarity of concepts is a cyclical (iterative) task to obtain the greatest semantic similarity).

The module for semantic similarity of concepts assessing works in interaction with external to the module blocks, namely: the repository of ontologies, external repositories of Big Data and metadata.

The development prospects of the proposed method are the formation of means of structuring NL descriptions using background knowledge of the structure of the task/query, but without SA knowledge, because only by performing the above comparison the user gains access to pertinent SA ontologies. If such a search is iterative, knowledge of the ontology that the user wants to replace with a more suitable one for the purpose can be used to search for more pertinent ontologies.

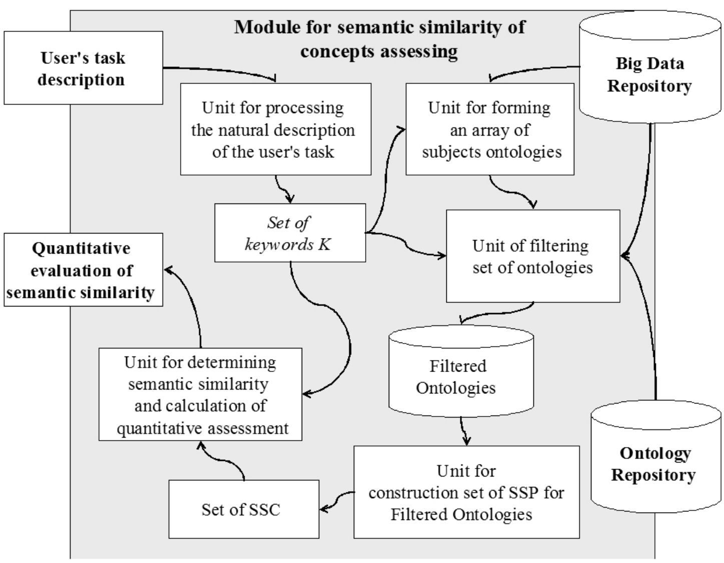

Fig. 3. Module for semantic similarity of concepts assessing for an intelligent cybersecurity system.

Proposed method implements the components of the theory of semantic recognition, interpretation, selection and structuring of data and allows creating so-called adaptive ontologies, which are most optimized for a specific applied task of artificial intelligence, and also allows more accurate structuring and selection of data for intelligent applications. The method involves designing ontology classes of a subject area or information object for the current situation (task) and introducing metrics within the ontology of a metric, with which to search for the required semantic distance.

CONCLUSIONS

The proposed approach to data integration and structuring in an intelligent information system is based on semantic analysis of metadata and semantic similarity determination of structural data models (ontologies, data), as well as the formation of a set of similar ontologies for solving problems of artificial intelligence.

The application of the created method of definition/evaluation of semantic similarity of concepts, which provides formation of an array of query features and researched ontologies, filtering of these features and closeness degree determination of researched ontology to user query characteristics, it is a tool for semantic comparison of ontologies found in repository under formal conditions with poorly structured natural language description.

Currently, there is no generally accepted standard for presenting metadata, so the proposed methods of analysis of NL annotations are the most adequate means of comparing the semantics of ontologies, data with those tasks for which they can be applied.

The scope of the developed method and module can be used in artificial intelligence systems for big data processing, cybersecurity, competence analysis when creating a team for the project implementation, human resource management, field of finance and business, companies working with dynamically-modified content of documents (jurisprudence, finance, standardization, public authorities etc.).

REFERENCES

1. Vasyukov V.L. Formal ontology and artificial intelligence. Moscow: IP RAS, 2006. 140 p. (In Russian)
2. Gladun A.Ya., Rogushina Yu.V. Ontology repositories as a means of reusing knowledge for recognizing information objects. *Design Ontology*.2013, No. 1 (7), pp. 35–50. (In Russian)
3. Torsten Hahmann Ontology repositories: A treasure trove for content ontology design patterns. *Proceedigs of 8th Inter. Workshop on Modular Ontologies* (WoMO-2014), Rio de Janeiro, Brazil, 2014.
4. Kenneth Baclawski, Todd Schneider. The open ontology repository initiative: Requirements and research challenges. *Proc. 8th International Semantic Web Conference*, ISWC 2009, Chantilly, VA, USA, 2010.
5. Rachel Heery and Sheila Anderson. Digital repositories review, February 2005.
6. William H. Inmon and W. H. Inmon. Building the Data Warehouse, 3rd Edition. Wiley, New York, 2002.
7. Jens Hartmann, York Sure, Raul Palma, Peter Haase, Mari Carmen Suarez-Figueroa, Rudi Studer, Asuncion Gomez-Perez. Ontology metadata vocabulary and applications. *Int Conf on Ontologies, Databases and Applications of Semantics. Workshop on Web Semantics (SWWS)*, Oct 2005.
8. Vekhorev M.N., Panteleev M.G. Construction of repositories of ontological knowledge bases. *Software products and systems*. 2011, No. 3 (3), pp.67–89.

9. Taylor C. An Introduction to Metadata. The University of Queensland, Australia. URL: http://www.libraty.uq.edu.au/papers/ctmeta4.html (Last accessed: 17.03.2018)
10. Corcho O. Ontology based document annotation: trends and open research problems. *Intern. Journal of Metadata, Semantics and Ontologies*. Vol. 1, Issue 1, January 2006. URL: http://www.dia.fi.upm.es/~ocorcho/documents/IJMSO2006_Corcho (Last accessed: 21.05.2020)
11. ISO 15489-1:2016 Information and documentation — Records management — Part 1: Concepts and principles.
12. ISO 15836-1:2017 Information and documentation — The Dublin Core metadata element set — Part 1: Core elements.
13. ISO 15836-2:2019 Information and documentation — The Dublin Core metadata element set — Part 2: DCMI Properties and classes.
14. DSTU ISO 15489-1: 2018 Information and documentation. Records management. Part 1. Concepts and principles (ISO 15489-1: 2016, IDT). (In Ukrainian)
15. DSTU ISO 15836-1: 2018 Information and documentation. Dublin Core Metadata Element Set. Part 1. Basic elements (ISO 15836-1: 2017, IDT). (In Ukrainian)
16. Gladun A., Khala K., Subach I. Ontological Approach to Big Data Analytics in Cybersecurity Domain. *Collection "Information Technology and Security"*. 2020, Vol. 8, No. 2, pp.120–132. DOI: https://doi.org/10.20535/2411-1031.2020.8.2.222559. URL: http://its.iszzi.kpi.ua/article/view/222559 (Last accessed: 11.12.2020)
17. Obrst, L., Ceusters, W., Mani, I., Ray, S., & Smith, B. (2007). The evaluation of ontologies. In Semantic web (pp. 139–158). Springer, Boston, MA.
18. Gómez-Pérez, A. (2004). Ontology evaluation. In Handbook on ontologies (pp. 251–273). Springer, Berlin, Heidelberg.
19. Leo Obrst, Werner Ceusters, Inderjeet Mani, Steve Ray, Barry Smith. The evaluation of ontologies. In Christopher J.O. Baker and Kei-Hoi Cheung, editors, *Revolutionizing Knowledge Discovery in the Life Sciences*, chapter 7. Springer, Berlin, 2007, pp. 139–158.
20. Obrst L. Ontologies for semantically interoperable systems. *Proceedings of the 12th International Conference on Information and knowledge management*, 2003, pp. 366–369.
21. Gangemi A., Catenacci C., Ciaramita M., Lehmann J. Modelling ontology evaluation and validation. *European Semantic Web Conference*. Springer, Berlin, Heidelberg, 2006, pp. 140–154.
22. Jens Hartmann, Raul Palma, Asunci´on Gomez-Perez "Ontology Repositories". Chapter in *Handbook on Ontologies*. Springer, 2009, pp 551–571.
23. Yimin Wang, Jie Bao, Peter Haase, and Guilin Qi. Evaluating formalisms for modular ontologies in distributed information systems. *Proc. of the First Inter Conf on Web Reasoning and Rule Systems* (RR2007), LNCS 4524, (June, 2007, Innsbruck). Innsbruck, Austria, pp. 178–182.
24. Jens Hartmann, Raul Palma. OMV – Ontology Metadata Vocabulary. *Semantic Web*. 2006, Vol. 2.0, URL: http://omv.ontoware.org/ (Last accessed: 17.12.2020)
25. Palma R., Hartmann J., Haase P. OMV Ontology Metadata Vocabulary. *SemanticWeb*. 2009, 76 p. URL: http://kent.dl.sourceforge.net/project/omv2/OMV%20Documentation/OMV-Reportv2.4.1.pdf. (Last accessed: 12.09.2017)
26. Jens Hartmann, Raúl Palma, Peter Haase, Asunción Gómez-Pérez. "Ontology Metadata Vocabulary and Applications." *OTM Workshops*, 2005, pp. 906–915.
27. Welcome to OMV — Ontology Metadata Vocabulary — http://omv2.sourceforge.net/. (Last accessed: 28.07.2020)
28. Gladun A.Y.,Khala K.A. Ontology-based semantic similarity to metadata analysis in the information security domain. *Prombles in programming*. 2021, №2, pp.34-41. DOI: htpps://doi.org/10.15407/pp2021.02.034. URL: http://pp.isofts.kiev.ua/ojs1/article/view/459 (Last accessed: 08.01.2021)
29. Tversky A. Features of Similarity. *Psychological Rev*. 1977, V. 84, P. 327.
30. Rogushina J. Use of Semantic Similarity Estimates for Unstructured Data Analysis. CEUR Vol-2577, *Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security"* (ITS 2019), Kyiv, 2019, pp. 246–258. URL: http://ceur-ws.org/Vol-2577/paper20.pdf. (Last accessed: 11.10.2019)
31. Gladun A.Y., Rogushina Y.V. Development of domain thesaurus as a set of ontology concepts with use of semantic similarity and elements of combinatorial optimization. *Prombles in*

programming. 2021, №2, pp. 3–15. DOI: htpps://doi.org/10.15407/pp2021.02.003. URL: http://pp.isofts.kiev.ua/ojs1/article/view/456 (Last accessed: 08.01.2021)

Received 11.01.2021

ЛІТЕРАТУРА

1. Васюков В.Л. Формальная онтология и искусственный интеллект. Москва: ИФ РАН, 2006. 140 с.
2. Гладун А.Я., Рогушина Ю.В. Репозитории онтологий как средство повторного использования знаний для распознавания информационных объектов // Онтология проектирования, № 1 (7), 2013. С.35–50.
3. Torsten Hahmann Ontology repositories: A treasure trove for content ontology design patterns. *In Proceedigs of 8th Inter. Workshop on Modular Ontologies* (WoMO–2014), Rio de Janeiro, Brazil, 2014.
4. Kenneth Baclawski Todd Schneider The open ontology repository initiative: Requirements and research challenges. In *Proc. 8th International Semantic Web Conference*, ISWC 2009, Chantilly, VA, USA, 2010.
5. Rachel Heery and Sheila Anderson. Digital repositories review, February 2005.
6. William H. Inmon and W. H. Inmon. Building the Data Warehouse, 3rd Edition. Wiley, New York, 2002.
7. Jens Hartmann, York Sure, Raul Palma, Peter Haase, Mari Carmen Suarez-Figueroa, Rudi Studer, Asuncion Gomez-Perez. Ontology metadata vocabulary and applications. *In Int Conf on Ontologies, Databases and Applications of Semantics.* In Workshop on Web Semantics (SWWS), Oct 2005.
8. Vekhorev M.N., Panteleev M.G. Construction of repositories of ontological knowledge bases. *Software products and systems*. 2011. No. 3 (3). P. 67–89.
9. Taylor C. An Introduction to Metadata. The University of Queensland, Australia. http://www.libraty.uq.edu.au/papers/ctmeta4.html
10. Corcho O. Ontology based document annotation: trends and open research problems. *Intern. Journal of Metadata, Semantics and Ontologies.* Volume 1. Issue 1. January 2006. URL: http://www.dia.fi.upm.es/~ocorcho/documents/IJMSO2006_Corcho (Дата звернення: 21.05.2020)
11. ISO 15489-1:2016 Information and documentation — Records management — Part 1: Concepts and principles.
12. ISO 15836-1:2017 Information and documentation — The Dublin Core metadata element set — Part 1: Core elements.
13. ISO 15836-2:2019 Information and documentation — The Dublin Core metadata element set — Part 2: DCMI Properties and classes.
14. ДСТУ ISO 15489-1:2018 Інформація та документація. Керування записами. Частина 1. Поняття та принципи (ISO 15489-1:2016, IDT).
15. ДСТУ ISO 15836-1:2018 Інформація та документація. Набір елементів метаданих Дублінського ядра. Частина 1. Основні елементи (ISO 15836-1:2017, IDT).
16. Gladun A., Khala K., Subach I. Ontological Approach to Big Data Analytics in Cybersecurity Domain . *Collection "Information Technology and Security"*. 2020. Vol. 8 No. 2. P.120–132. DOI: https://doi.org/10.20535/2411-1031.2020.8.2.222559. URL: http://its.iszzi.kpi.ua/article/view/222559 (Дата звернення:11.12.2020)
17. Obrst, L., Ceusters, W., Mani, I., Ray, S., & Smith, B. (2007). The evaluation of ontologies. In Semantic web (pp. 139–158). Springer, Boston, MA.
18. Gómez-Pérez, A. Ontology evaluation. *In Handbook on ontologies*. Springer, Berlin, Heidelberg, 2004. P. 251–273.
19. Leo Obrst, Werner Ceusters, Inderjeet Mani, Steve Ray, Barry Smith. The evaluation of ontologies. In Christopher J.O. Baker and Kei-Hoi Cheung, editors, *Revolutionizing Knowledge Discovery in the Life Sciences*. Chapter 7. 2007. P. 139–158.
20. Obrst L. (2003, November). Ontologies for semantically interoperable systems. *In Proceedings of the twelfth international conference on Information and knowledge management* (pp. 366–369).

21. Gangemi, A., Catenacci, C., Ciaramita, M., & Lehmann, J. (2006, June). Modelling ontology evaluation and validation. *In European Semantic Web Conference*. Springer, Berlin, Heidelberg. P. 140–154.
22. Jens Hartmann, Raul Palma, and Asunci´on Gomez-Perez "Ontology Repositories". *Chapter in Handbook on Ontologies*. Springer, 2009. P. 551–571.
23. Yimin Wang, Jie Bao, Peter Haase, Guilin Qi. Evaluating formalisms for modular ontologies in distributed information systems. *In Proc of the First Inter Conf on Web Reasoning and Rule Systems* (RR2007), LNCS 4524, Innsbruck, Austria, Jun 2007. P. 178–182.
24. Jens Hartmann and Raul Palma. OMV — Ontology Metadata Vocabulary for the Semantic Web, 2006. v. 2.0, URL: http://omv.ontoware.org/. (Дата звернення: 17.12.2020)
25. Palma R., Hartmann J., Haase P. OMV Ontology Metadata Vocabulary for the SemanticWeb. 2009. 76p.URL: http://kent.dl.sourceforge.net/project/omv2/ OMV%20Documentation/OMV-Reportv2.4.1.pdf. (Дата звернення: 12.09.2017)
26. Jens Hartmann, Raúl Palma, Peter Haase, Asunción Gómez-Pérez. Ontology Metadata Vocabulary and Applications. *OTM Workshops*, 2005, P. 906–915.
27. Welcome to OMV – Ontology Metadata Vocabulary. URL: http://omv2.sourceforge.net/. (Дата звернення: 28.07.2020)
28. Gladun A.Y., Khala K.A. Ontology-based semantic similarity to metadata analysis in the information security domain. *Prombles in programming*. 2021. №2. P. 34–41. DOI: htpps://doi.org/10.15407/pp2021.02.034. URL: http://pp.isofts.kiev.ua/ojs1/article/view/459 (Дата звернення:08.01.2021)
29. Tversky A. Features of Similarity. *Psychological Rev.* 1977. V. 84. P. 327.
30. Rogushina J. Use of Semantic Similarity Estimates for Unstructured Data Analysis // CEUR Vol-2577, *Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security"* (ITS 2019), Kyiv, 2019. P. 246–258. URL: http://ceur-ws.org/Vol-2577/paper20.pdf. (Дата звернення: 11.10.2019)
31. Gladun A.Y., Rogushina Y.V. Development of domain thesaurus as a set of ontology concepts with use of semantic similarity and elements of combinatorial optimization. *Prombles in programming*. 2021. №2. P. 03–15. DOI: htpps://doi.org/10.15407/pp2021.02.003. URL: http://pp.isofts.kiev.ua/ojs1/article/view/456 (Дата звернення: 08.01.2021)

Отримано 11.01.2021

Гриценко В.І.[1], член-кореспондент НАН України,
директор Міжнародного науково-навчального центру
інформаційних технологій та систем НАН України та МОН України
ORCID: 0000-0003-4813-6153, e-mail: vig@irtc.org.ua
Гладун А.Я.[1], канд.техн.наук, старш. наук. співроб.,
відд. комплексних досліджень інформаційних технологій та систем
ORCID: 0000-0002-4133-8169, e-mail: glanat@yahoo.com
Хала К.О.[1],
наук. співроб., відд. комплексних досліджень інформаційних технологій та систем
ORCID: 0000-0002-9477-970X, e-mail: cecerongreat@ukr.net
Мартінез-Бежар Р.[2], д-р. філософії (інформатика),
професор каф. інформаційно-комунікаційних технологій та штучного інтелекту
ORCID: 0000-0002-9677-7396, e-mail: rodrigo@um.es
[1] Міжнародний науково-навчальний центр інформаційних
технологій та систем НАН України та МОН України,
пр. Академіка Глушкова, 40, м. Київ, 03187, Україна
[2] Університет Мурсії,
CP 30180 Буллас, Іспанія

МЕТОД ОЦІНЮВАННЯ СЕМАНТИЧНОЇ БЛИЗЬКОСТІ ПОНЯТЬ ДЛЯ СПІВСТАВЛЕННЯ ОНТОЛОГІЙ В ПРИКЛАДНИХ ЗАДАЧАХ ШТУЧНОГО ІНТЕЛЕКТУ

Вступ. Розвиток інтелектуальних інформаційних технологій (ІІТ) передбачає створення нового класу систем на основі формалізованого подання знань про предметну область.

Нині в різних предметних областях вже створено велику кількість онтологій. Сучасні засоби дають змогу шукати серед них бажану онтологію лише за деякими формальними параметрами (наприклад, за ключовими словами), а не на рівні їхньої семантики. Вирішення проблем пошуку необхідної онтології може бути забезпечено сховищем онтологій, яке уможливлює оброблення знань онтологій.

Мета статті — розроблення алгоритмів та методів оцінювання семантичних моделей, що полягають у поєднані якісного (онтологічного) подання знань з кількісним (числовим) оцінюванням онтологій та їхніх параметрів (семантична близькість, семантична відстань, семантична спорідненість), що спрямовано на віднайдення подібності між елементами різних онтологій.

Методи. Для розв'язання поставлених завдань використані методи онтологічного аналізу об'єктів предметної області, теоретико-множинні підходи до визначення міри близькості двох об'єктів шляхом зіставлення їхніх властивостей (feature matching) та традиційні методи статистичного аналізу.

Результати. Запропонований метод оцінювання семантичної подібності дає змогу на основі семантичного аналізу природномовних анотацій метаданих як онтологій, так і даних (зокрема великих даних), уможливлює виконання завдання їхньої інтерпретації та відбору. Отримані результати надають можливості створення оригінальних інтелектуальних інформаційних систем для економіки, медицини, національної безпеки, оборони та соціальної сфери.

Висновки. Запропонований оригінальний підхід до оцінювання та аналізу метаданих (онтологій, даних) базується на семантичному аналізі метаданих та визначенні семантичної подібності структурних моделей даних (онтологій, даних) і формуванні ранжованого набору близьких онтологій для розв'язання завдань штучного інтелекту. Застосування методів визначення семантично близьких понять подано як інструмент для семантичного зіставлення структури онтологій, які було знайдено у репозиторії за формальними умовами, зі слабо структурованим описом природною мовою (ПМ). Запропоновані методи аналізу ПМ-анотацій є адекватним засобом зіставлення семантики онтологій, даних з тими задачами, для розв'язання яких вони можуть застосовуватися.

Ключові слова: *метадані, стандарти метаданих, семантична подібність, формальна модель онтології, інтелектуальна інформаційна система, репозиторій онтології.*