UDC 004.896:004.912:004.048 https://doi.org/10.15407/pp2022.03-04.355

SEMANTIC ALIGNMENT OF ONTOLOGIES MEANINGFUL CATEGORIES WITH THE GENERALIZATION OF DESCRIPTIVE STRUCTURES

Eduard Manziuk, Olexander Barmak, Iurii Krak,
Olexander Pasichnyk, Pavlo Radiuk, Olexander Mazurets

The presented work addresses the issue of semantic alignment of ontology components with a generalized structured corpus. The field of research refers to the sphere of determining the features of trust in artificial intelligence. An alignment method is proposed at the level of semantic components of the general alignment system. The method is a component of a broader alignment system and compares entities at the level of meaningful correspondence. Moreover, only the alignment entities' descriptive content is considered within the proposed technique. Descriptive contents can be represented by variously named id and semantic relations. The method defines a fundamental ontology and a specific alignment structure. Semantic correspondence in the form of information scope is formed from the alignment structure. In this way, an entity is formed on the side of the alignment structure, which would correspond in the best meaningful way to the entity from the ontology in terms of meaningful descriptiveness. Meaningful descriptiveness is the filling of information scope. Information scopes are formed as a final form of generalization and can consist of entities, a set of entities, and their partial union. In turn, entities are a generalization of properties that are located at a lower level of the hierarchy and, in turn, are a combination of descriptors. Descriptors are a fundamental element of generalization that represent principal content. Descriptors can define atomic content within a knowledge base and represent only a particular aspect of the content. Thus, the element of meaningfulness is not self-sufficient and can manifest as separate meaningfulness in the form of a property, as a minimal representation of the meaningfulness of an alignment. Descriptors can also supplement the content at the level of information frameworks, entities, and properties. The essence of the alignment in the form of information scope cannot be represented as a descriptor or their combination. It happens because the descriptive descriptor does not represent the content in the completed form of the correspondence unit. The minimum structure of representation of information scope is in the form of properties. This form of organization of establishing the correspondence of the semantic level of alignment allows you to structure and formalize the information content for areas with a complex form of semantic mapping. The hierarchical representation of the generalization not only allows simplifying the formalization of semantic alignment but also enables the formation of information entities with the possibility of discretization of content at the level of descriptors. In turn, descriptors can expand meaningfulness at an arbitrary level of the generalization hierarchy. This provides quantization of informational content and flexibility of the alignment system with discretization at the level of descriptors. The proposed method is used to formalize the semantic alignment of ontology entities and areas of structured representation of information.

Keywords: semantic alignment, ontology, information scope, entities

В статті розглядається проблема семантичного порівняння складових онтології з узагальненим структурованим корпусом. Область дослідження відноситься сфери визначення складових довіри до штучного інтелекту. Пропонується метод порівняння на рівні семантичних складових загальної системи порівняння. Метод є складовою більш широкої системи порівняння та здійснює порівняння сутностей на рівні змістовної відповідності. В пропонованому методі враховуються тільки описові змістовності сутностей порівняння. Описові змістовності можуть бути представленні різними іменованими назвами та структурними зв'язками. В методі визначається базова онтологія та певна структура порівняння. Із структури порівняння формується семантична відповідність у вигляді інформаційних рамок. Таким чином на стороні структури порівняння формується сутність, яка б найкращим змістовним чином відповідала сутності з онтології за змістовною описовістю. Змістовна описовість є наповненням інформаційних рамок. Інформаційні рамки формується у вигляді кінцевої форми узагальнення та можуть складатися із сутностей, сукупності сутностей та їхнього часткового об'єднання. В свою чергу сутності є узагальненням властивостей, які розташовані на нижчому рівні ієрархії та свою чергу є поєднанням описових дескрипторів. Описові дескриптори є базовим елементом узагальнення та є представленням базової змістовності. Описові дескриптори можуть визначати атомарну змістовність в межах бази знань та представляють лише певний аспект змістовності. Таким аспект змістовності не є самодостатнім та можу проявлятись у виді відокремленої змістовності у вигляді властивості, як мінімальні формі представлення змістовності порівняння. Також описові дескрипторі можуть доповнювати змістовність на рівні інформаційних рамок, сутностей та властивостей. Сутність порівняння у вигляді інформаційних рамок не може бути представлена у вигляді описового дескриптора або їх поєднанням. Це зумовлено тим, що описових дескриптор не представляє змістовність в завершеній формі одиниці відповідності. Мінімальна форма представлення інформаційних рамок полягає у вигляді властивостей. Така форма організації встановлення відповідності семантичного рівня порівняння дозволяє структурувати та формалізувати інформаційну змістовність для областей із складною формою семантичного відображення. Ієрархічне представлення узагальнення дозволяє не тільки спростити формалізацію семантичного порівняння, а також дає змогу формувати інформаційні сутності із можливістю дискретизації змістовності на рівні описових дескрипторів. В свою чергу описові дескрипторі можуть розширити змістовність на довільному рівні ієрархії узагальнення. Це забезпечує квантування інформаційної змістовності та гнучкість системи порівняння з дискретизаціє на рівні описових дескрипторів. Запропонований метод використовується для формалізації семантичного порівняння сутностей онтології та областей структурованого представлення інформації.

Ключові слова: семантичне вирівнювання, онтологія, інформаційна рамка, сутності

Introduction

The need to establish the correspondence of concepts in different subject areas arises in connection with the simultaneous and rapid development of applied research areas. Accordingly, a set of the diversity of formulations and content representations is formed under such circumstances. The variety of such forms can also appear for reasons unrelated to the specifics of the subject area. Since there is a wide field of similar research or those with

© У.Ф. Манзюк, О.В. Бармак, Ю.В. Крак, О.А. Пасічник, П.М. Радюк, О.В. Мазурець, 2022

identical research essences in terms of meaningful comparison, there is a need to establish the correspondence of the meaningful essences of the research subject areas. The need to develop the degree of entity alignment within subject areas is topical both in cybersecurity [27, 28] and in other subject areas [7, 14, 20, 21, 26, 37], including in machine learning methods [4–6, 19, 25]. Such an alignment is challenging due to the ambiguity of the interpretation and the existing objective circumstance, which consists of incomplete correspondence. That is, there is a certain inconsistency in the content of the entities. This discrepancy creates uncertainty that significantly affects the level of compliance. In some circumstances, the existing non-conformity does not play a significant role; in other cases, the same level of non-conformity can have a decisive effect. Determining the level of inconsistency is essential when establishing the consistency of the entities of the alignment areas. However, there is a problem with formalization, especially in the case of comparing the content of concepts that have the form of entities. In this regard, there is a need to develop alignment methods that allow formalizing the establishment of correspondences relative to a particular basic set of entities.

The primary subject area of the alignment is determined, in which the essence of the alignment is defined. The next stage is finding a representation of meaningful correspondence in the comparison field. Accordingly, there is a need to define the basic unit of a certain amount of content. The following sections present the definition of correspondences according to this descriptive content.

Related works

To determine the semantic alignment, let's take several studies that are the closest in determining correspondence and comparing entities from subject areas. They set the limits of informativeness in the form of a core, which can be expanded by supplementing it with specific properties, using inversion, symmetry, intersection, union, and other forms [11]. Although there are known structural alignment methods [9, 23], the semantic approach focuses on the search for common meaningfulness. The semantic correspondence between gene ontology terms determined from gene annotations and used within bioinformatics is determined [40]. The definition of similarity between entities is defined based on finding a weighted path between concepts and finding distances between them [33]. It is essential to formalize information from the subject area for further comparison [41], which is also manifested in forming comparison entities for the field of electronic commerce when comparing goods [18]. The problem of establishing correspondence also arises when comparing ontologies from the standpoint of multilingualism [15], and this necessity is also manifested with the emergence of the Semantic Web [10]. The problem of knowledge alignment arises when comparing knowledge bases that are represented in different languages [35].

At the same time, the alignment problem is relevant in comparing large ontologies, which requires the use of automated methods for dividing ontologies into smaller areas of alignment [16] and comparing large knowledge bases [34]. The alignment of large ontologies is carried out by automated clustering methods with subsequent alignment [30]. The computerized knowledge alignment systems are being developed but require human correction to improve quality alignment [32]. However, it remains basic to determine the alignment quality based on experts' assessments [2]. To improve the automated alignment, machine learning methods are used [8, 24], including when using neural networks [13] and random forest classifiers [31]. However, the quality of the automatic alignment largely depends on the quality of the data [1]. Important is research towards ontology-based knowledge alignment, in which entities are aligned based on their position in alignment ontologies [40].

Semantic knowledge alignment methods are most often used as a more straightforward form of knowledge representation and alignment [37]. In this case, the alignment task is greatly facilitated since knowledge acquires a more formalized representation. However, this requires pre-processing. Furthermore, the ambiguity problem when comparing entities remains relevant when using knowledge bases [12, 42], which manifests itself when searching for text similarity of content [36]. Another form of addition is the introduction of additional external information to fill the significant gap between ontologies [3], which can be presented in the form of general-purpose background knowledge [31]. The corpus is formed on many resources that independent parties developed with differences in the representation of the same phenomenon in the real world. These resources are developed in pursuit of different goals and from a wide range of applied areas, and the developers are geographically distant, affecting the seen and presented an image of the real world in the research subject area.

According to the analysis, there is a significant part of the research on comparing ontologies entities that concern particular areas and solve specific problems, for example, in the field of bioinformatics or construction. At the same time, automatic methods have insufficient comparison quality and must be refined with a human's help. Yet, means and techniques that would help a person to standardize semantic comparison are insufficiently developed. Therefore, there is an urgent need to establish a formalized approach to the discretization of the representation of informativeness when establishing correspondences of the essences of ontologies.

Proposed semantic alignment

The basis of the developed method for the applied problem of comparison is an approach using knowledge bases in the form of ontologies or schemes. This approach is also used for the comparison of ontologies, generalization, etc. Taking this method as a basis, we will improve it for a specific case within the framework of the study of the correspondence of the trust ontology to AI and the structured domain based on the gray literature corpus of the subject area. The heterogeneity of knowledge-based data is a special case of the more general concept of diversity. Diversity, in general, generates incompleteness of perception and causes the integration of knowledge that is presented in various forms of external representation in order to achieve the maximum possible perception that corresponds to the realities of the surrounding world. Diversity is widespread in the description of the surrounding world. For the same phenomenon, observers will give different descriptions regarding meaningfulness, objectivity, necessity, and other factors. This is based on different views, experiences, goals, language, etc., which is manifested in the complexity of knowledge integration.

When knowledge is formally represented in the form of a corpus of texts, semantic heterogeneity will be presented in the form of the diversity of language $DivLang$ and diversity of knowledge $DivKnow$ representation. Let's present the general diversity when comparing within the research as the diversity of informativeness $DivInf = DivKnow \cup DivLang$. Varieties of the language are manifested in linguistic phenomena (for example, synonyms and homonyms for another).

The diversity of informativeness is manifested in the assumed absence of linguistic diversity in the diversity of content of the entity and is key in semantic heterogeneity. The content of entities can be overdetermined and accumulate the properties of less general entities, which is manifested by the absorption of other entities by the content, and partial overlapping of the content of entities. To obtain a set of knowledge within a certain concept, which is presented in the form of a named entity, let's introduce a more general concept that summarizes entities at the meta-level and represents a certain reference entity in relation to comparison tasks. Let's denote the goal level of entities as $ScIn$ (Scope Information). The purpose of information frameworks is to generalize knowledge, which is presented in the form of a set of heterogeneous entities and is actually a semantic representation of one and the same element of the surrounding world. At the level of the concept, the presence of heterogeneity in the presentation of information in the sources is assumed. Heterogeneity becomes the basis of the method and is not a problem since it practically manifests itself within the boundaries of the corpus and, therefore, is known. Thus, the heterogeneity is limited by the corpus. The information framework is initially unknown, although it is defined by knowledge and limited by the corpus.

We define the formal model of the information framework as follows

$$ScIn = \langle Ent, Prop(Des), Scp \rangle. \tag{1}$$

Here Ent – a set of named entities; $Prop(Des)$ – a set of properties that form named entities and are determined by a set of descriptions; Scp – the range of properties is defined by a tuple.

$$Scp = \{\langle ent, Scp(ent)\rangle \mid \forall ent \in Ent, ent\ є\ сутністю\ ScIn\} \tag{2}$$

If defined, the framework of properties in relation to the entity

$$Scp(ent) = \{\forall prop \in Prop \mid prop\ є\ властивістю\ ent\} \tag{3}$$

The diversity of language $DivLang$ is formalized in the presence of a set of named entities. Named entities can formalize the same knowledge by a set of properties but have different names.

The diversity of knowledge $DivKnow$ is formalized by quantizing knowledge in the form of properties $Prop(Des)$. The same properties, for example, can form different named entities forming a non-empty intersection of properties belonging to certain entities.

We will present the diversity of the language in the form of such concepts as synonyms and hyperonyms.

Linguistic correspondence at the level of synonyms

$$\begin{gathered} match_{synonym}(ent, ent') = synonym(ent, ent') \\ s.t.\ \ synonym(ent, ent') \equiv synonym(ent', ent) \end{gathered} \tag{4}$$

To find synonyms, i.e., entities that have different nominal designations and a fully coincident set of property descriptions defined $\{(dsc_{ent}, dsc_{ent'}) \mid \forall dsc_{ent} \in Dsc_O, \forall dsc_{ent'} \in Dsc_{Ds}, \forall dsc(dsc_{ent} \equiv dsc_{ent'})\}$.

Linguistic correspondence at the level of hyperonyms

$$\begin{gathered} match_{hyperonym}(ent, ent') = hyperonym(ent, ent') \\ s.t.\ \ hyperonym(ent, ent') \not\equiv hyperonym(ent', ent) \end{gathered} \tag{5}$$

To find hyperonyms, that is, entities, the relationship between which is that one entity is more general than the other. Taken in relation to descriptions, the entity that is a hyperonym is above the plural relative to another entity. defined in this way $\{dsc_{ent} \mid \forall ent \in Ent_O\} \subset \{dsc_{ent'} \mid \forall ent' \in Ent_{Ds}\}$, the inverse relation is also defined $\{dsc_{ent'} \mid \forall ent' \in Ent_{Ds}\} \subset \{dsc_{ent} \mid \forall ent \in Ent_O\}$.

Within the framework of the structured domain, linguistic heterogeneity is defined by codes relating synonyms and hyperonyms to ethical entities. Such diversity is limited to diversity within named designations. Further research is conducted within the framework of the diversity of knowledge representation. However, given that the knowledge representation within the corpus is linguistic, establishing linguistic correspondence at the description level dsc is an important step and is performed for each description. The most important and desirable thing is to establish the descriptions according to the correspondence of the synonyms of the descriptions $match_{synonym}$ because such correspondence is maximal. Hyperonym matching $match_{hyperonym}$ is the next step if no synonym matches are found. The matching of hyperonyms requires the establishment of a threshold value for taking into account the matching and measure of content. If the measure of content is too small, that is, the hyperonym too generalizes a description from another domain of comparison, such a relation is considered inappropriate.

Semantic alignment is defined by the function $semantic(\ )$. Next, we will define a descriptive definition of the matching method.

A quantized unit of knowledge is obtained based on descriptions *Des* from the Knowledge Base. Knowledge Base is limited to the body of documents and is formed by it. In the diversity of knowledge, certain knowledge is formed by separate descriptions and their aggregates and is formally presented in the form of properties *Prop*(*Des*).

Thus, the central place in semantic heterogeneity is occupied by knowledge, which can have a different form of representation, both structural and linguistic, but unchanged in the sense of informativeness. The body of knowledge is grouped in the form of a certain association. The association is grouped formally in the form of informative frameworks, which represent a certain conceptual structure, the purpose of which, in the case of semantic comparison, is to correspond to the essence of the basic set of the ontology of trust. Note that in the basic set, the entity itself is compared as a fixed and formed element. This entity is formed by the existing informative framework of compliance on the side of the structured domain. That is, the correspondence is determined by the set of knowledge presented in the form of formalized properties within the information scope. The unit of comparison is the informative frame, that is, the unit of the meta-level above the entities of the structured domain. Let's depict a graphic representation of the formation of an information scope element (Figure 1).

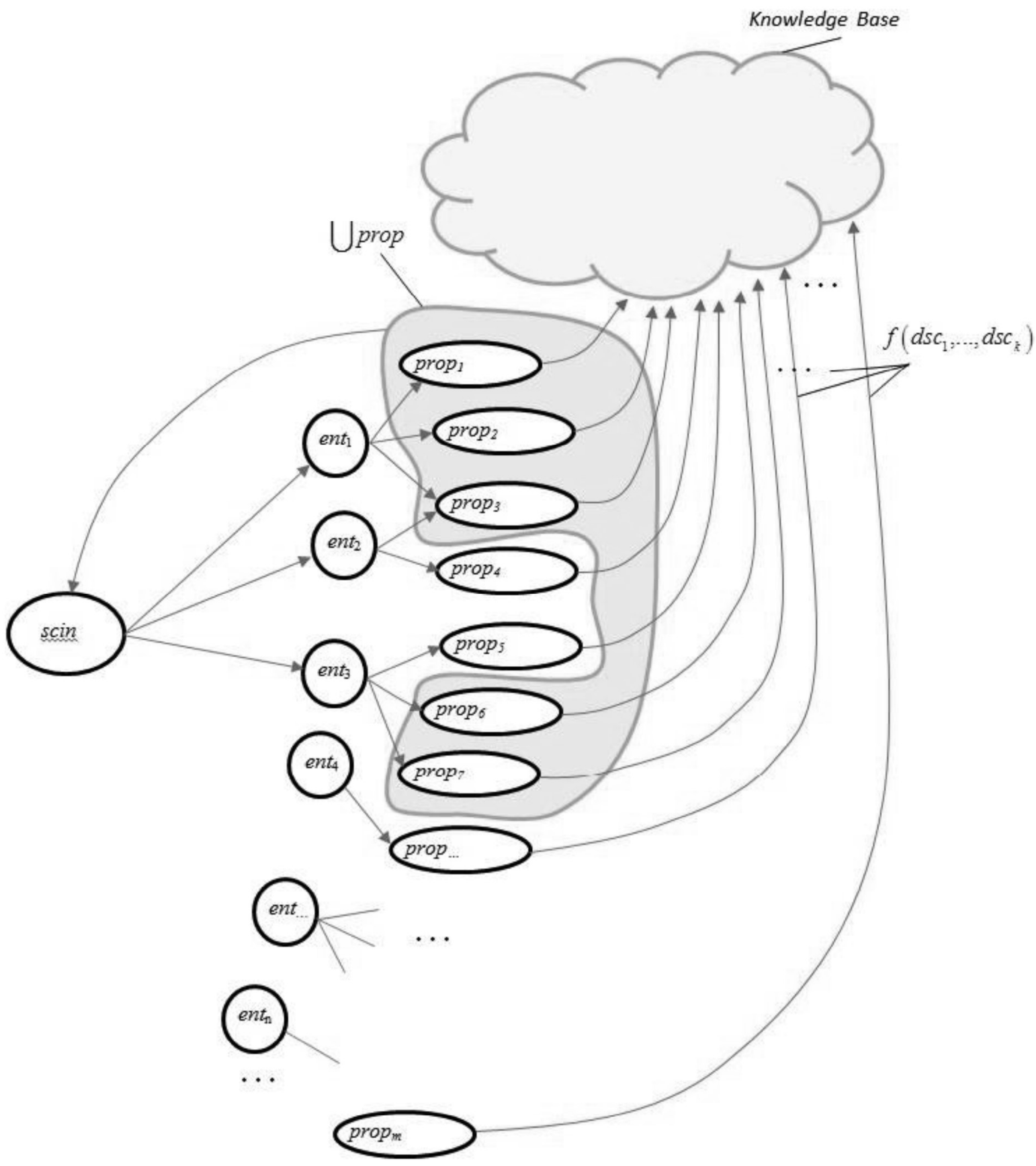

Figure 1. Formation of an element of an information scope *ScIn* in the form of aggregate properties as a function of descriptions $Prop(Dsc)$ obtained from the knowledge base

Properties are formed from the knowledge base by extracting individual descriptions. A set of properties that can be generalized by common informativeness forms an entity and is a subset of properties $\bigcup prop_{ent_i} \subset Prop$. A set of entities that can be summarized by common informativeness at the entity level form an informative scope and are a subset of entities $\bigcup ent_{scin_i} \subset Ent$. Joint informativeness at the level of entities in relation to an information scope can be bound by a partial relation, since it may not fully belong to the feature of generalization and refer to an information scope. That is, not all properties that form a certain entity can be included in the information scope.

Thus, the information scope is a certain meta entity that can, in some cases, summarize several defined entities in the domain *Ds* or partially correlate with other entities at the level of properties; however, it is possible to match the research entity from the domain as much as possible in terms of the degree of correspondence *O*. This corresponds to the main purpose of its formation according to the semantic method of comparison.

An entity is not a simple generalization or formal naming of a subset of properties. An entity is a generalization of knowledge and informativeness over a subset of properties, considering the relationships between entities. The same ratio connects the concepts of essence and information scope. At the same time, at the generalization level, entities can broadcast the generalization of knowledge and informativeness to the information scope through the generalization of properties. This relationship structure is depicted in Figure 2.

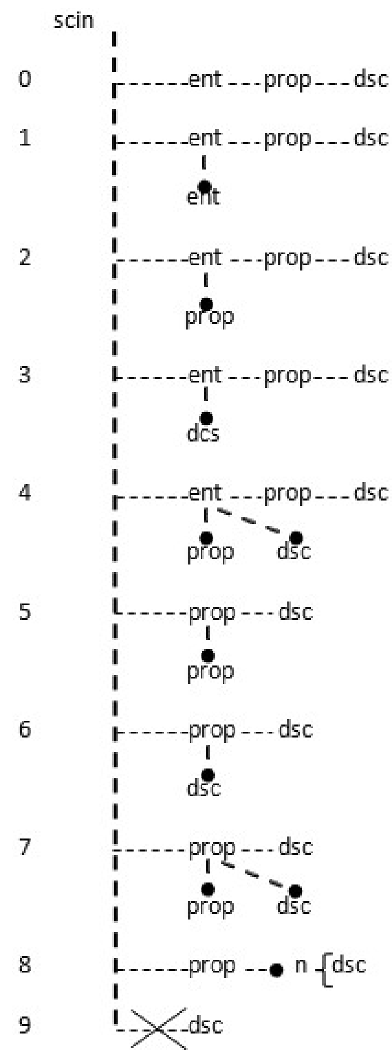

Figure 2. The main variants of the presentation form of the information scope

Thus, an information scope is used for comparison, as a generalization of knowledge at the level of properties $scin_{Ds} \Rightarrow ent_O$. The set of properties is summarized by an information scope in such a way as to correspond to a certain entity from the ontology of trust as much as possible. This method isolates knowledge to determine the degree of alignment.

This allows for a flexible alignment method. The need for the presented model of establishing the degree of correspondence is justified by the fact that the named entities on the set of the structured domain do not always and do not fully correspond to the entities from the ontology according to meaningful criteria. Accordingly, to ensure a qualitative comparison and considering the semantic heterogeneity of corpus documents, a meta-entity is formed on the side of the structured domain, the main purpose of which is to maximally correspond to the entity on the ontology side in terms of content.

Thus, the method is also justified by the fact that the structured domain is less formalized than the ontology and has greater heterogeneity of both language and knowledge representation. This gives more opportunities for searching and forming generalizations for correspondences, and if necessary and insufficient informativeness of the structured domain, the original source of information from the corpus of documents is obtained. The decision to establish compliance is formed using the descriptions of the corpus documents. The description of the relevant property is taken from a certain set of documents to ensure objectivity and diversity of views.

The main variants of the form of presentation of the informative frame, shown in Figure 2, can be used in combination. Option 9 is not implemented because the descriptor is not a complete representation of the content. The totality of the set of descriptors $\{dcs\}_1^n$ that form property $prop = \{dcs\}_1^n$ is represented by option 8. This is the minimal form of formation of an informative framework with the definition of content $scin_{\min} : scin = prop = \{dcs\}_1^n$. Other options allow expansion when an informative frame is formed by expanding the basic formation.

Thus, we define correspondence for semantic alignment as follows

$$semantic(ent, ent' \Rightarrow scin) = \frac{\left|\sigma_{prop_{Ds} \in Prop_{Ds} \wedge \left(Fun_{semanticPropertie}(prop_O, prop_{Ds}) > threshhold\right)}\left(\pi_{ent}\left(Prop_O\right)\right)\right|}{\left|\pi_{ent}\left(Prop_O\right)\right|} \tag{6}$$

The semantic alignment function is defined using the semantic alignment of properties and is a function of the set of descriptions dsc that are obtained from the ontology domain $dom\ O$ and the structured domain of the corpus $dom\ Ds$, respectively.

$$semanticPropertie(prop_O, prop_{Ds}) = f\left(\{dsc_O\}, \{dsc_{Ds}\}\right) \tag{7}$$

The number of found alignment properties on the set of ontology properties relative to the essence of the research $\pi_{ent}(Prop_O)$ and limited by the selection function from those properties that have a match on the set Ds, i.e. determined $semanticPropertie(prop_O, prop_{Ds})$ whose value is greater than the *threshold*, is estimated. The assessment is determined by the total number of the study essence properties ent. To establish a unified designation of the entities of the syntactic, structural, and semantic levels, a certain entity ent' from Ds. This entity can be considered as a first approximation of the corresponding entity from Ds. In the future, the entity can move into an information scope $ent' \Rightarrow scin$, the set of properties of which corresponds to a greater extent to the set of properties of the research entity by the heterogeneity of language and knowledge representation.

Experimental studies

Experimental studies were conducted to determine the effectiveness of the method of semantic determination of the alignment of ontologies content categories with the generalization of descriptive structures. Alignment is performed for an ontology [24] and a certain structured corpus [17].

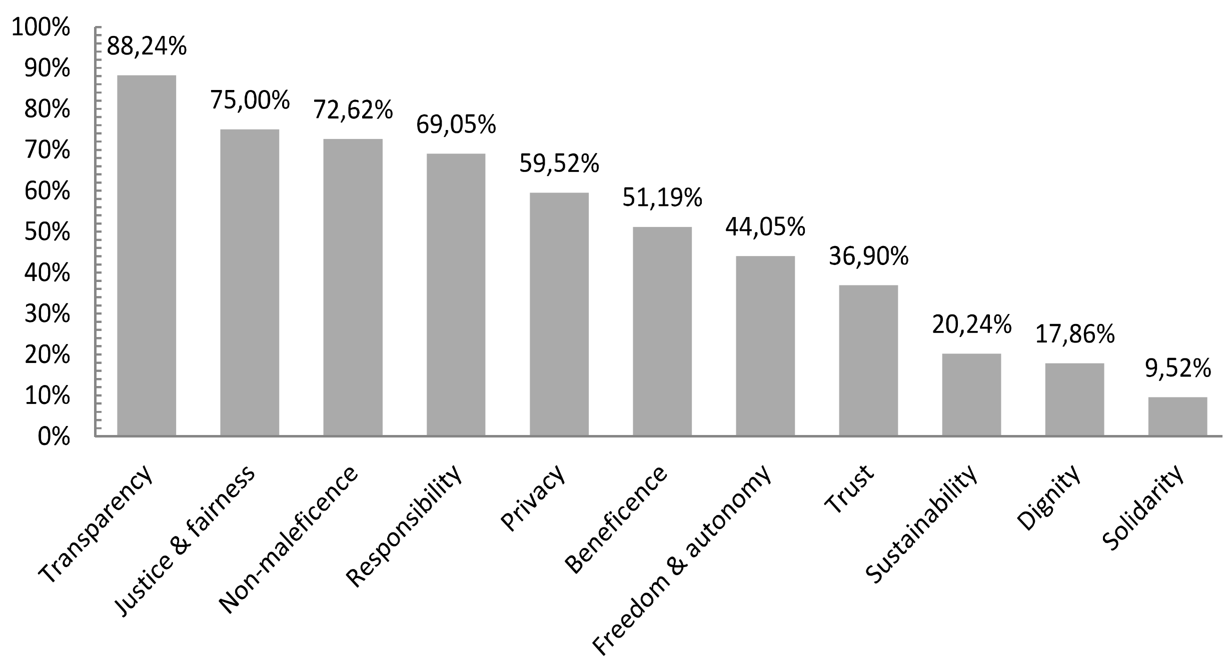

Figure 3. Distribution of AI trust ontology concepts within the corpus

Analysis using semantic comparison of ontologies and a structured domain allows for determining the relevance of the importance of AI reliability components. There are variations in the names of concepts; that is, concepts with a similar structure can have different lexical names. However, when using semantic comparison, the names of the concepts do not play an important role. In the conceptual categories {Transparency} and {Privacy}, their prevalence and importance is determined by the shares of 88.24% and 59.52%, respectively. According to the content complexity, the content of {Explainability} is embedded in {Transparency}. Further, the comparison using lexical variability using the content definition {strategies for reduction Bias} corresponds to the concept {Justice & fairness}, and {functional Safety} corresponds to the concept {Non-maleficence} and has 75% and 72.62% importance. According to the smaller corresponding fraction, the partial semantic match {Controllability} has a lexical counterpart {Freedom & autonomy} with a prevalence of 44.05%. {Trust} concept with a spread of 36.9%. Concepts {Responsibility} are presented in the composition of 69.05%, {Beneficence} respectively 51.19%. Concepts {Sustainability}, {Dignity}, {Solidarity} are presented at the level of group closeness. The purpose of the comparison is to include concepts and generalizations as much as possible. The comparison demonstrates the effectiveness of the semantic comparison of ontology and structured domains.

Conclusions and discussion

The developed method of alignment of the essences of ontologies or structured domains allows formalizing the alignment process. It provides significant diversity in the form of representation of informativeness. Such cases arise due to the non-strict correspondence of the essences of the alignment, significant inconsistency, and descriptive complexity. The method allows determining the basic content unit as a descriptor. This allowed the field of knowledge to be presented in a discrete form with the form of an atomic representation of content. Further hierarchical generalization makes it possible to form properties that form an informative frame as entities of maximum correspondence to the object of alignment. The method provides the discretization of the knowledge domain and the necessary flexibility for a formal approach. The form of presentation of knowledge in the area of matching with the application of the proposed method does not play a significant role. This is because the entities of correspondence are formed and built from a discrete field of atomic representation of meaningfulness. Moreover, meaningfulness is formed in separate knowledge and not in the form of dispersed descriptiveness of unique content. The application of the method of semantic determination of the correspondence of meaningful categories of ontologies with the generalization of descriptive structures allows for formalizing the alignment process and improving the quality of semantic inference.

References

1. Algergawy A., Cheatham M., Faria D., Ferrara A., Fundulaki I., Harrow I., Hertling S., Jiménez-Ruiz E., Karam N., Khiat A., Lambrix P., Li H., Montanelli S., Paulheim H., Pesquita C., Saveta T., Schmidt D., Shvaiko P., Splendiani A., Thiéblin E., Trojahn dos Santos C., Vatascinov J., Zamazal O., Zhou L. Results of the Ontology Alignment Evaluation Initiative 2018: *13th International Workshop on Ontology Matching co-located with the 17th ISWC (OM 2018)*, Monterey, United States , October 2018. Pp.76–116.
2. Althobaiti A. F. S. Comparison of Ontology-Based Semantic-Similarity Measures in the Biomedical Text. *Journal of Computer and Communications*. 2017. Vol. 05, No. 02. Pp. 1–17. URL: https://doi.org/10.4236/jcc.2017.52003.
3. Annane A., Bellahsene Z. GBKOM: A generic framework for BK-based ontology matching. *Journal of Web Semantics*. 2020. Vol. 63. Pp. 100563. URL: https://doi.org/10.1016/j.websem.2020.100563.
4. Barmak O., Krak Y., Manziuk E. Characteristics for choice of models in the ansables classification. *CEUR Workshop Proceedings*. 2018. Vol. 2139. Pp.171–179. URL: https://doi.org/10.15407/pp2018.02.171.
5. Barmak O., Krak I., Manziuk E. Diversity as The Basis for Effective Clustering-Based Classification. *CEUR-WS*. 2020. Vol. 2711. Pp. 53–67.
6. Barmak O., Manziuk E., Krak I. Using piecewise hyper linear classification in multidimensional feature space for text content: *2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv, Ukraine , September 17, 2019. Pp.119–123. URL: https://doi.org/10.1109/STC-CSIT.2019.8929798.
7. Boiko J., Pyatin I., Eromenko O., Stepanov M. Method of the adaptive decoding of self-orthogonal codes in telecommunication. *Indonesian Journal of Electrical Engineering and Computer Science*. 2020. Vol. 19, No. 3. Pp. 1287–1296. URL: https://doi.org/10.11591/ijeecs.v19.i3.pp1287-1296.
8. Chen J., Jiménez-Ruiz E., Horrocks I., Antonyrajah D., Hadian A., Lee J. Augmenting Ontology Alignment by Semantic Embedding and Distant Supervision: *The Semantic Web*, Cham , Springer International Publishing, 2021. Pp.392–408. URL: https://doi.org/10.1007/978-3-030-77385-4_23.
9. Chu S.-C., Xue X., Pan J.-S., Wu X. Optimizing Ontology Alignment in Vector Space. *Journal of Internet Technology*. 2020. Vol. 21, No. 1. Pp. 15–22.
10. Drakopoulos G., Voutos Y., Mylonas P. Recent Advances On Ontology Similarity Metrics: A Survey: *2020 5th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*, September 2020. Pp.1–7. URL: https://doi.org/10.1109/SEEDA-CECNSM49515.2020.9221837.
11. Garanina N., Sidorova E., Kononenko I., Gorlatch S. Using Multiple Semantic Measures for Coreference Resolution in Ontology Population. *International Journal of Computing*. 2017. Vol. 16, No. 3. Pp. 166–175.
12. Hao J., Chen M., Yu W., Sun Y., Wang W. Universal Representation Learning of Knowledge Bases by Jointly Embedding Instances and Ontological Concepts: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, NY, USA , Association for Computing Machinery, July 25, 2019. Pp.1709–1719. URL: https://doi.org/10.1145/3292500.3330838.
13. Holter O. M., Myklebust E. B., Chen J., Jimenez-Ruiz E. Embedding OWL ontologies with OWL2Vec. *CEUR Workshop Proceedings*. 2019. Vol. 2456. Pp. 33–36.
14. Hrytsyk V., Nazarkevych M. Real-Time Sensing, Reasoning and Adaptation for Computer Vision Systems: *Lecture Notes in Computational Intelligence and Decision Making*, Cham , Springer International Publishing, 2022. Pp.573–585. URL: https://doi.org/10.1007/978-3-030-82014-5_39.
15. Ivanova T., Popov M. Ontology Evaluation and Multilingualism: *Proceedings of the 21st International Conference on Computer Systems and Technologies '20*, New York, NY, USA, Association for Computing Machinery, June 19, 2020. Pp.215–222. URL: https://doi.org/10.1145/3407982.3407989.
16. Jiménez-Ruiz E., Agibetov A., Chen J., Samwald M., Cross V. Dividing the Ontology Alignment Task with Semantic Embeddings and Logic-based Modules. *arXiv:2003.05370 [cs]*. 2020.
17. Jobin A., Ienca M., Vayena E. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*. 2019. Vol. 1, No. 9. Pp. 389–399. URL: https://doi.org/10.1038/s42256-019-0088-2.
18. Kakad S., Dhage S. Ontology Construction from Cross Domain Customer Reviews using Expectation Maximization and Semantic Similarity: *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, March 2021. Pp.19–23. URL: https://doi.org/10.1109/ESCI50559.2021.9396780.

19. Krak I., Barmak O., Manziuk E. Approach to Piecewise-Linear Classification in a Multi-dimensional Space of Features Based on Plane Visualization. *Computational Intelligence*, 2022, 38(3), pp. 921–946. DOI:10.1111/coin.12289.
20. Krak, Y. V., Barmak, A. V., Baraban, E. M. Usage of NURBS-approximation for construction of spatial model of human face. *Journal of Automation and Information Sciences*, 2011. 43(2), 71-81. doi:10.1615/JAutomatInfScien.v43.i2.70
21. Kryvonos I. G., Krak I. V., Barmak O. V., Bagriy R. O. New tools of alternative communication for persons with verbal communication disorders. *Cybernetics and Systems Analysis*, 2016. 52(5), 665-673. doi:10.1007/s10559-016-9869-3
22. Lastra-Díaz J. J., Goikoetxea J., Hadj Taieb M. A., Garcia-Serrano A., Ben Aouicha M., Agirre E., Sánchez D. A large reproducible benchmark of ontology-based methods and word embeddings for word similarity. *Information Systems*. 2021. Vol. 96. Pp. 101636. URL: https://doi.org/10.1016/j.is.2020.101636.
23. Manziuk E., Krak I., Barmak O., Mazurets O., Kuznetsov V., Pylypiak O. Structural alignment method of conceptual categories of ontology and formalized domain. *CEUR-WS*. 2021. Vol. 3003. Pp. 11–22.
24. Manziuk E., Barmak O., Krak I., Mazurets O., Skrypnyk T. Formal Model of Trustworthy Artificial Intelligence Based on Standardization. *CEUR-WS*. 2021. Vol. 2853. Pp. 190–197.
25. Manziuk E. A., Wójcik W., Barmak O. V., Krak I. V., Kulias A. I., Drabovska V. A., Puhach V. M., Sundetov S., Mussabekova A. Approach to creating an ensemble on a hierarchy of clusters using model decisions correlation. *Przegląd Elektrotechniczny*. 2020. Vol. 96, No. 9. Pp. 108–113. URL: https://doi.org/10.15199/48.2020.09.23.
26. Martsenyuk V., Bernas M., Klos-Witkowska A., Gancarczy T. On Parallel Processing of Machine Learning Based On Big Data and Voronoi Tessellation: *CEUR Workshop Proceedings*, 2021. Pp.104–113.
27. Morozova O., Nicheporuk A., Tetskyi A., Tkachov V. Methods and technologies for ensuring cybersecurity of industrial and web-oriented systems and networks. *Radioelectronic and computer systems*. 2021. No. 4. Pp. 145–156. URL: https://doi.org/10.32620/reks.2021.4.12.
28. Nicheporuk A., Savenko O., Nicheporuk A., Nicheporuk Y. An Android Malware Detection Method Based on CNN Mixed-data Model2020. Pp.198–213.
29. Nkisi-Orji I., Wiratunga N., Massie S., Hui K.-Y., Heaven R. Ontology Alignment Based on Word Embedding and Random Forest Classification: *Machine Learning and Knowledge Discovery in Databases*, Cham , Springer International Publishing, 2019. Pp.557–572. URL: https://doi.org/10.1007/978-3-030-10925-7_34.
30. Patel A., Jain S. A Partition Based Framework for Large Scale Ontology Matching. *Recent Patents on Engineering*. 2020. Vol. 14, No. 3. Pp. 488–501. URL: https://doi.org/10.2174/1872212113666190211141415.
31. Portisch J. P. Towards Matching of Domain-Specific Schemas Using General-Purpose External Background Knowledge: *The Semantic Web: ESWC 2020 Satellite Events*, Cham , Springer International Publishing, 2020. Pp.270–279. URL: https://doi.org/10.1007/978-3-030-62327-2_42.
32. Qi Z., Zhang Z., Chen J., Chen X., Zheng Y. PRASEMap: A Probabilistic Reasoning and Semantic Embedding based Knowledge Graph Alignment System. *arXiv:2106.08801 [cs]*. 2021.
33. Rathee P., Malik S. K. IWD towards Semantic similarity measure in ontology. *Journal of Information and Optimization Sciences*. 2020. Vol. 41, No. 7. Pp. 1561–1577. URL: https://doi.org/10.1080/02522667.2020.1802129.
34. Rinaldi A. M., Russo C., Madani K. A Semantic Matching Strategy for Very Large Knowledge Bases Integration. *International Journal of Information Technology and Web Engineering (IJITWE)*. 2020. Vol. 15, No. 2. Pp. 1–29. URL: https://doi.org/10.4018/IJITWE.2020040101.
35. Singh H., Jain P., Mausam, Chakrabarti S. Multilingual Knowledge Graph Completion with Joint Relation and Entity Alignment. *arXiv:2104.08804 [cs]*. 2021.
36. Sunilkumar P., Shaji A. P. A Survey on Semantic Similarity: *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*, December 2019. Pp.1–8. URL: https://doi.org/10.1109/ICAC347590.2019.9036843.
37. Vedernikov M., Zelena M., Volianska-Savchuk L., Litinska V., Boiko J. Management of the Social Package Structure at Industrial Enterprises on the Basis of Cluster Analysis. *TEM Journal*. 2020. Vol. 9, No. 1. Pp. 249-260. URL: https://doi.org/10.18421/TEM91-35.
38. Xiang Y., Zhang Z., Chen J., Chen X., Lin Z., Zheng Y. OntoEA: Ontology-guided Entity Alignment via Joint Knowledge Graph Embedding. *arXiv:2105.07688 [cs]*. 2021.
39. Zeng K., Li C., Hou L., Li J., Feng L. A comprehensive survey of entity alignment for knowledge graphs. *AI Open*. 2021. Vol. 2. Pp. 1–13. URL: https://doi.org/10.1016/j.aiopen.2021.02.002.
40. Zhao C., Wang Z. GOGO: An improved algorithm to measure the semantic similarity between gene ontology terms. *Scientific Reports*. 2018. Vol. 8, No. 1. Pp. 1–10. URL: https://doi.org/10.1038/s41598-018-33219-y.
41. Zhong B., Li H., Luo H., Zhou J., Fang W., Xing X. Ontology-Based Semantic Modeling of Knowledge in Construction: Classification and Identification of Hazards Implied in Images. *Journal of Construction Engineering and Management*. 2020. Vol. 146, No. 4. Pp. 04020013. URL: https://doi.org/10.1061/(ASCE)CO.1943-7862.0001767.
42. Zhu G., Iglesias C. A. Exploiting semantic similarity for named entity disambiguation in knowledge graphs. *Expert Systems with Applications*. 2018. Vol. 101. Pp. 8–24. URL: https://doi.org/10.1016/j.eswa.2018.02.011.

Received 03.08.2022

Про авторів:

[1]*Манзюк Едуард Андрійович,*
доцент кафедри комп'ютерних наук;
86 друк. праць, в т.ч.: 48 наук. статей у фахових вид.,
25 зарубіжних публікацій
Індекс Хірша - 5
ORCID iD: 0000-0002-7310-2126

[2]*Бармак Олександр Володимирович,*
доктор технічних наук, професор,
завідувач кафедри комп'ютерних наук ;
понад 100 друк. праць, в т.ч.: 45 наук. статей у фахових вид.,
12 зарубіжних публікацій
Індекс Хірша - 10
ORCID iD: 0000-0003-0739-9678

[3]*Крак Юрій Васильович,*
доктор фізико-математичних наук, професор,
завідувач кафедри теоретичної кібернетики;
понад 700 друк. праць, в т.ч.: 177 наук. статей у фахових вид.,
93 зарубіжних публікацій;
Індекс Хірша - 14
ORCID iD: 0000-0002-8043-0785

[4]*Пасічник Олександр Анатолійович*,
доцент кафедри комп'ютерних наук ;
122 друк. праць, в т.ч.: 49 наук. статей у фахових вид.,
1 зарубіжних публікацій
Індекс Хірша - 1
ORCID iD: 0000-0002-8760-4688

[5]*Радюк Павло Михайлович,*
PhD; старший викладач кафедри комп'ютерних наук
20 друк. праць, в т.ч.: 16 наук. статей у фахових вид.,
10 зарубіжних публікацій
Індекс Хірша - 2
ORCID iD: 0000-0003-3609-112X

[6]*Мазурець Олександр Вікторович*
кандидат технічних наук, доцент кафедри комп'ютерних наук
155 друк. праць, в т.ч.: 31 наук. статей у фахових вид.,
11 зарубіжних публікацій
ORCID iD: 0000-0002-8900-0650

Місце роботи авторів:

[3]Київський національний університет імені Тараса Шевченка,
01601, Київ, вул. Володимирська, 60;

Інститут кібернетики імені В.М.Глушкова НАН України,
03187, Київ, пр. Глушкова, 40;
E-mail: Iurii.krak@knu.ua, yuri.krak@gmail.com

[1,2,4,5,6]Хмельницький національний університет МОН України,
29016, Хмельницький, вул. Інститутська, 11;
E-mail: eduard.em.km@gmail.com, alexander.barmak@gmail.com,
o.a.pasichnyk@gmail.com, radiukpavlo@gmail.com, exechong@gmail.com

Прізвища та ініціали авторів і назва доповіді англійською мовою:
Manziuk E.A., Barmak O.V., Krak Iu.V., Pasichnyk O.A., Radiuk P.M., Mazurets O.V.
Semantic alignment of ontologies meaningful categories with the generalization
of descriptive structures

Прізвища та ініціали авторів і назва доповіді українською мовою:
Манзюк Е.А., Бармак О.В., Крак Ю.В., Пасічник О.А., Радюк П.М., Мазурець О.В.
Метод семантичного визначення відповідності змістовних категорій онтологій
із узагальненням описових структур