Kostiantyn Tkachenko,
PhD of Economical Sciences, AssociatePprofessor,
Associate Professor at the Department
of Computer SystemsSsoftware Department,
National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»,
Kyiv, Ukraine
tkachenko.kostyantyn@gmail.com
https://orcid.org/0000-0003-0549-3396

SEMANTIC ANALYSIS OF NATURAL LANGUAGE TEXTS: ONTOLOGICAL APPROACH

The development of information (intelligent) learning systems, electronic document management systems, web-oriented systems working with text information in natural language has led to an increase in the volume of educational content and/or arrays of processed full-text documents. All this requires new means of organizing access to information, many of which should be classified as intelligent systems for knowledge processing. One of the effective approaches to identifying and processing the meaning of educational content (and/or text documents) is the use of ontologies.

The purpose of the article is research, analysis of various approaches to determining the semantic content of texts in natural language, consideration of existing concepts of text analysis and prospects for using the proposed ontological approach to semantic analysis of texts in natural language

Research methods are methods of semantic analysis of the main concepts of the analyzed subject area (semantic analysis of texts in natural language). The article considers an approach to the linguistic analysis of texts based on ontological modeling.

The novelty of the research is the application of the proposed ontological approach to the semantic analysis of texts in natural language to determine the meaning (semantics) of text information, which is used in intelligent systems of various classes.

The conclusion of the research carried out in the article is as follows: an ontological approach to the semantic analysis of natural language text, its tasks and methods is proposed. The use of the proposed approach to text analysis leads to the understanding of semantic analysis as

a single triad: <Ontology – Text – Meaning>. For effective and correct extraction of knowledge, it is suggested to use a multi-level ontology. The result of the interaction of a specific natural language text with an ontology is an ontological meaning – a set of interconnected subgraphs of the ontograph.

The ontological content is extracted from the ontograph using a semantic analyzer. The dialogue processor examines the syntactic tree of sentence parsing (a connected element of the natural language text) and, based on the given question, finds a fragment in the text that is the answer to the question. Computer understanding (in information or intellectual systems) of natural language text is achieved, in particular, by: immersion of the text in a single environment of knowledge – ontology; formal presentation of meaning (semantics) in the knowledge base of the corresponding system; the possibility of operations on the ontological content. The proposed approach can be used to create intelligent information repositories that work in a single knowledge environment.

The proposed approach to the semantic analysis of texts in natural language is focused on the automatic extraction of metadata from texts of various nature (for example, a text document, Internet content, educational content of relevant online courses, description of computer and board games). With further development of the proposed approach, it can be used in systems of automatic referencing of scientific publications, meaningful interpretation of multimedia content, training and testing (including elements of visual display of information and elements of gamification).

Keywords: natural language text; linguistic analysis of natural language text; semantic analysis of natural language text; ontology; subject area; knowledge base, ontograph.

Introduction. The development of information (intelligent) learning systems, electronic document management systems, web-oriented systems working with text information in natural language has led to an increase in the volume of educational (training) content and/or arrays of processed full-text documents (Atzeni and Atzori, 2018; Maulud et al., 2021).

All this requires new means of organizing access to information, many of which should be classified as intelligent systems for knowledge processing (Tkachenko, 2024; Li, 2020; Tkachenko et al., 2020). One of the effective approaches to identifying and processing the meaning of educational content (and/or text documents) is the use of ontologies (Semantic Analysis, n.d.).

The problems of linguistic analysis of natural language texts to determine their meaning (semantics) and the development of corresponding ontologies, algorithms for linguistic (including syntactic and semantic) analysis of texts require their solution, and their relevance is beyond doubt.

To solve these problems, it is proposed to form an ontology of the subject area (SA), reflecting the collection of fragments of the natural language text (for example, some text document, Internet content, educational content of relevant online courses, descriptions of computer and board games (text documents), and/or full-text documents), representing natural language texts.

The formed ontology contributes to solving current problems of linguistic (in particular, semantic) analysis of natural language texts, in particular, determining the meaning (semantics) of the analyzed texts.

The work examines the essence of linguistic (in particular, semantic) analysis of natural language texts.

The main goal of the article is research, analysis of different approaches to determining the semantic content of texts in natural language, consideration of existing concepts of text analysis and prospects for using the proposed ontological approach to semantic analysis of texts in natural language.

Natural language texts can be, for example, text documents, Internet content, educational content of relevant online courses, descriptions of computer and board games.

The tasks of the research, in particular, are:

– analysis of modern approaches to linguistic analysis of texts in natural language;

– consideration of the essence of semantic analysis of texts in natural language;

– the use of an ontological approach to determining the semantics of texts in the semantic analysis of texts, which is based, in particular, on the fact that:

• each text (text fragment) is a set of terms of the concepts of the corresponding SA;

• the set of fragments of texts in natural language is divided into subsets of texts that are close in content (so-called clusters), thereby obtaining groups of terms of concepts of the same topic;

• establishment of relationships between terms and concepts (each term is characterized by its own weight (priority). Terms with a weight greater than the average determine the terms of the ontology, and the terms of the concepts with the maximum weight are chosen as concepts.

Methods of linguistic analysis of natural language texts are diverse and depend on the problem being solved in the problem domain.

The following can be distinguished:

– methods of syntactic analysis of natural language texts;

– methods of morphological analysis of natural language texts;

– methods of semantic processing of texts, which are aimed at:

• "linguistic transformations" (for example: translation into a foreign language and back; summary; note-taking; thesis presentation; annotation);

• "knowledge extraction" (for example: message classification, answering questions, contextual translation and understanding of discourses (Sowa, 2002)) based on the use of artificial intelligence (AI) methods, in particular, conceptual analysis methods. In this case, the following problems can be distinguished: 1) synthesis of knowledge representation systems – ontologies; 2) development of systems of semantic analysis and machine "understanding" of texts using ontologies.

The first problem is solved by different approaches, for example:

– automatic extraction of knowledge substructure in the problem domain from the corpus of professional NL texts, when formal-logical and syntactic analysis tools are used to synthesize the ontology (Palagin, 2008);

– development of knowledge structures with the help of experts and knowledge engineers, when textbooks, recent publications and other manuals on describing the problem domain serve as a ready-made base for ontology developers (Mochalova and Mochalov, 2016; Fu et al. 2014; Eremin, 2008; Gavrilova and Voinov, 2007).

The approach to solving the second problem is as follows (Gladun et al., 2009):

If the *description* of the situation presented in the text can be achieved by purely linguistic means, then *understanding* the situation is possible beyond the linguistic resource of the text – by mobilizing cognitive efforts and individual knowledge of a person. A coherent perception of the text is possible only through its understanding (Lanin, 2015; Sanfilippo, 2018; List, 2018).

Research results. Understanding people is based on matching external information with their cognitive model of the world, and «computer understanding» is based on displaying information in a specific and formally defined system of knowledge. When machine processing a natural language text, it is necessary to find a projection of the text onto a computer ontology.

Ontology defines terms used to describe and represent knowledge of SA. Ontologies include definitions of the main concepts of the SA and the relationships between them that are accessible for computer processing (Gelfert, 2017).

Formally, ontology O_{SA} can be defined as a set (Tkachenko et al., 2020)

$$OSA = <DL, C, FL, FC, Rh>,$$

where $D_L = \{(T_i, r_i)\}$ (i = 1, . . . , n) is a dictionary of terms of problems of concepts;

T_i is a term (possibly for several words), r_i is the rating of the term T_i relative to other terms of the concept.

C is a set of concepts, $C = \{C_i\}$ (i = 1, . . . , m).

$F_L(L) \rightarrow C$ is the function of interpreting terms T_i of the problem under consideration. F_L associates a set of terms from the dictionary D_L with a certain subset of concepts.

$F_C(C_i) \rightarrow D_L$ is the function of interpreting concepts, which matches the concepts C_i with a set of terms from the dictionary D_L.

R_h is a set of the hierarchical relations between concepts from the set C.

The specific task of AI is as follows. The natural language text is given, its semantic analysis ("understanding") implies, in particular, the definition of: topic, content, message intention, communicative meaning.

Machine understanding of natural language text is an AI paradigm. To obtain the formal (computer) meaning of natural language texts, it is necessary to conduct the corresponding semantic analysis of these texts.

For such an analysis, it is necessary to consider a system of semantic analysis of natural language texts, which should provide for and combine the following basic procedures:

– grammatical analysis;
– interaction of the text with the ontology,
– obtaining the result – formal understanding of the text through the ontology;
– definition/clarification of the meaning of the text.

The proposed technology of linguistic (especially semantic) analysis implies achieving better mutual understanding between the author of the text and its consumer through a certain intelligent system, a common knowledge base and language.

Natural language text is a linguistic, informational and cultural phenomenon that is relevant for a given period of existence of society and can be in demand by descendants.

It follows that the text material is initially "immersed" in a certain universal system of accumulation and interpretation of knowledge, in which it itself was generated.

The text is generated both individually and collectively and can be claimed individually or collectively. This means that the content ("semantic content") of the text is often multifaceted (multilevel), and each level has its own depth of presentation.

Performing the communication function, the text should be displayed in the user's (knowledge consumer's) knowledge base and interact with it. We will call the concentrated expression of knowledge the meaning of the natural language text.

The meaning must be extracted in order to work with it, store and cherish it, since it contains a certain author's intent, an intellectual resource and is a continuation of the corresponding natural language text. Extracting the meaning is not a one-time communication operation.

It is possible to fully or almost fully reveal the content resource of the natural language text only by repeatedly accessing it. This is the active function of the natural language text, which is used for its deep semantic analysis.

The system that performs semantic analysis of natural language texts must meet the following requirements:

− The partners of intellectual communication, together with the text, are immersed in a single computer environment of ontological knowledge.

− Linguistic processing of the source text (morphological, syntactic and semantic analysis of sentences) is necessary to identify terms that carry the content load.

− The result of computer semantic analysis of a coherent natural language text must be a formal/adapted text expressing its semantic content.

− The system must ensure control over the degree of adequacy of expression of thoughts expressed by the author of the natural language text (in some cases, the author may be the system itself).

− The system must be able to repeatedly activate the text for a more accurate definition of the meaning of the natural language text.

Thus, the potential capabilities of the natural language text are realized using analysis through ontology and/or active dialogue.

Providing functions for semantic analysis of natural language text within the framework of the corresponding information system presupposes, in particular:

− mutual understanding of communication partners, which is ensured by a unified system for representing knowledge accumulated in society; the contextual environment of communication can be a hierarchical ontology, which can be expanded and supplemented by a multitude of already existing subject-oriented ontologies;

− preliminary parsing of the text performed by the so-called linguistic analyzer, focused on the semantic analysis of ordinary text information (for example, in the linguistic model "Meaning-Text" there are several levels of expression representation, including syntactic, morphological and semantic. The linguistic analyzer must perform a whole range of actions, among which the following should be highlighted: building a syntactic parsing tree, identifying the main (so-called "core") sentence constructions of the natural language text, building relationships, defining "significant/essential/important" lexical groups (for example, keywords of the natural language text);

– extracting the meaning from the text, which is formalized through ontology – the set of subgraphs of the conceptual graph;

– identifying the meaning in the natural language text is performed by the semantic analyzer; 7

– correction of the natural language text by the author, who can compare the result of automatic meaning extraction with his understanding.

– the presence of a dialogue mode, which is implemented in the natural language by a dialogue processor, when the meaning can change, which is a reason for adjusting the ontology.

In the system that performs semantic analysis of natural language text, ontologies play a leading role. Ontology, as a system for displaying a common knowledge space (some authors call this space "general intelligence" (Lytvyn et al., 2018; Hermann et al., 2015)), allows interpreting text information in the language of common knowledge and unites the trinity <author – text – consumer> into a single intellectual environment.

The system that performs semantic analysis of natural language text is based on a multi-level ontology that has the following features:

1) knowledge of several levels is combined within the single structure: the highest level of abstraction, the middle level – public knowledge and the lower level – professional knowledge;

2) work with specific natural language texts;

3) the result of knowledge extraction from the text is the so-called "ontological meaning", which can be formalized and then processed in the corresponding information (or intelligent) system.

Multi-level ontology – conceptual ontological graph (ontograph) – the way of semantic (at the level of concepts used) representation of situations and knowledge in models of understanding the natural language, decision-making, reasoning, etc.

The nodes of the ontograph are lexical units – words of the natural language, which are considered as categories and concepts.

Categories and concepts are concepts that act as elements of the so-called «conceptual language» in lexical semantics.

Ontology can be considered as a pyramid of conceptual knowledge, in which concepts have varying degrees of generalization. The most abstract categories form the upper level of ontology. The concepts of the middle level form the descriptive continuum of knowledge (for example, knowledge base of SA), revealing the meanings of the upper-level categories through more frequently used vocabulary.

At the lower level are:

1) concepts that denote the concepts of everyday life, familiar objects and situations;

2) concepts that are related to professional knowledge of SA (for example, in the SA "Modeling a knowledge system" – this is a concept, relation, ontology, model, ontograph).

The nodes of the ontograph are connected by associative links. The orientation of the links in the graph is directed from concepts of the higher level of generalization

down to the concepts that characterize them. The concept of "the meaning of the situation" is interpreted differently in different disciplines.

The meaning is related to the goals of communication, is understood (explained) as the "structure of the situation", is associated with the category of "understanding" of the natural language text. The meaning can be described by a set of denotates freed from emotional, modal, pragmatic, stylistic and other shades.

Automation of the processes of semantic analysis of natural language texts stimulates the development of computational linguistics (CL) and computer models (for example, semantic networks, graph and hierarchical models, frame models, neural networks and, of course, ontological models). CL is the branch of linguistics in which the information technologies and computer science are applied to the analysis (in particular, lexical, semantic, syntactic) and synthesis of natural language.

Methods of CL are mainly reduced to morphological interpretation, operations on syntax and semntics. The result of semantic analysis of the natural language text is its formalized representation, the so-called "formalized meaning".

If we analyze the natural language text using an ontology, which is the knowledge structure corresponding to the problem of SA (domain), then the result will be concentrated knowledge that correlates with the text. The conceptual filter of the ontology yields the «conceptual (ontological) meaning».

An elementary meaning is defined with the help of the pair of ontograph nodes. Links can only record the fact of interaction between two words (for example, bee – insect, passenger – train, enterprise – industry).

An ontograph consists of the set of elementary meanings related to each other, which enter into permitted combinations. The connected part of the ontograph, connecting two nodes, forms the subgraph; when the directions of links (arrows) in it change to the opposite (from bottom to top), the subgraph chain is formed.

The chain of related elementary meanings, which begins at some "active" node and ends at the top of the ontology, forms an ontological chain of the active node. The chain, allocated by the active node on the ontograph, is called the ontological meaning of the active word.

The process of detecting the ontological meaning of the active (input) word begins with the allocation of the so-called "keyword" in the sentence of the natural language text. If such the word is present in the ontograph, then it makes the neighboring concept active and this continues until the exit to the higher levels of the ontology.

The result of such the process is an ordered sequence of related concepts, which determines the formal ontological meaning of the input word in the model under consideration for the specific SA. For example, the following keywords are allocated in the sentence: *bee* and *nectar*.

The corresponding conceptual chains will be constructed for them:
1) bee – Insect – Flight – Movement –Life;
2) nectar – Food – Life.

The connections in such a representation do not yet have an ontological meaning, therefore it is necessary to fix only the connection of the pair of objects.

Further, if the keywords are in the same sentence, then a new connection *bee -> nectar* will be automatically constructed, which was absent in the ontology.

As the result, the linear format of ontological meaning looks like this:

bee (insect, flight, movement, life) –> **nectar** (food, life.

Links can be interpreted by repeated reference to the text (the bee **collects** nectar).

Depending on the goals of semantic analysis, chains can be shortened by using abstract categories. Chains can be linear or branched. The result of the complete review of the natural language text is the set of semantic trajectories, which can be considered as the so-called "semantic portrait of the text".

Ontological meaning can be the goal and result of semantic analysis of the natural language text, because, in particular:

– keywords in the semantic chain are extracted from the natural language text;

– keywords are placed in the context of general knowledge, which is organized as an ontology;

– the set of semantic chains gives a brief formalized description of the natural language text (the text fragment of educational content or e-document) – the "semantic portrait text" in terms of general knowledge.

The ontological meaning extracted by information (or intelligent) system from the natural language text (for example, the certain text document, Internet content, educational content of relevant online courses, descriptions of computer and board games) becomes an element of the knowledge base, which is available to all communication partners.

Having formally defined the ontological meaning, we can talk about the so-called "computer understanding of the natural language text", which implies, in particular:

– identifying the so-called "author's intent" (with the help of an information system that uses the corresponding ontology);

– interpreting the "author's intent" within the knowledge system of the information (or intelligent) system itself;

– performing certain actions on the meaning, in particular, providing it with the natural language grammatical interpretation (often in the language of the consumer of knowledge and meanings).

Such computer understanding requires the presence of the following triple:

*<Ontol*ogy – Text – Meaning>.

The graphical structures of the trio support the process of understanding the text, as they help to obtain answers to the main questions of understanding, for example:

– "What is common between software engineering and computer science?";

– "What is the difference between a game of entertainment and a game of learning?".

In this case, the answers are given at the conceptual level of general knowledge – in the metalanguage of ontology.

The ontological meaning, which is a product of the semantic analysis of the natural language text, is realized in the form of trajectories that make the transition to an active state with the help of keywords selected in the natural language text.

The trajectories are ordered by the time of their appearance, they are semantically interconnected by natural language text fragments (one phrase, paragraph, section, etc.). Some keywords begin to interact only through common concepts of higher level.

Among all the trajectories, there are those that intersect and partially merge, and their concepts can intersect in different combinations and with different frequencies.

The value of identifying the ontological meaning is that it helps to obtain at least the most general and schematic, but quite adequate understanding of the essence of the natural language text.

If we use the approach considered in the simplified form to organize the dialogue, then the linguistic analyzer should be modernized.

When constructing a syntactic parsing tree of the sentence, the connections between the members of the sentence (arrows between tree nodes) should be put in correspondence with questions (indexed with questions), for example, such as: *who*, *who does what*, *what does*, *where*, *when*, *which*, *how much*, *why*, *how*, etc.

Such indexing creates an appropriate linguistic base for the dialogue.

Questions to the text, formulated by the user, activate the corresponding groups of words that form the answer and thereby reveal the deep meaning of the communication situation, using the established ontological meaning of the natural language text.

Conclusions. The proposed approach to semantic analysis (one of the types of linguistic) of natural language text, to the tasks of its semantic analysis, methods and result of processing natural language text leads to a single triad: <Ontology – Text – Meaning>.

The main task of semantic analysis of natural language text is to extract concentrated knowledge relevant to the author's intention, adequately reflecting it with the help of corresponding natural language texts.

For effective and correct extraction of knowledge, the work proposes to use a multi-level ontology – the conceptual system for representing general knowledge about both the world and the problem area under consideration.

The result of the interaction of specific natural language text with ontology is ontological meaning is the set of interconnected subgraphs of the ontograph. The ontological meaning is extracted from the ontograph using the so-called semantic analyzer.

For a more in-depth research of the content of the natural language text (for example, some text document, Internet content, educational content of relevant online courses, descriptions of computer and board games), the dialogue processor is used, which examines the syntactic parsing tree of the sentence (the coherent element of the natural language text) and, based on the given question, finds a fragment in the text that serves as a specific answer to the user's question.

Computer understanding (in information or intelligent systems) of the natural language text is achieved, in particular, due to:

1) immersion of the text in a single knowledge environment – ontology;

2) formal representation of the meaning in the knowledge base of the corresponding information (or intelligent) system;

3) the possibility of operations on the ontological meaning.

The proposed approach can be used to form databases, archive electronic documents, index them, classify them and search on the Internet. In addition, it can be used as the basis for creating intelligent information repositories (not only databases, but also knowledge base) operating in a single knowledge environment.

The proposed approach to linguistic, namely semantic, analysis of natural language texts is focused on the automatic extraction of metadata from natural language texts of various natures (for example, some text document, Internet content, educational content of relevant online courses, descriptions of computer and board games (text documents)).

With further development of the proposed approach, it can be used in systems, for example, automatic abstracting of scientific publications, meaningful interpretation of multimedia content, training and testing (including with elements of visual display of information and elements of gamification).

REFERENCES

Atzeni, M., Atzori, M., 2018. Translating natural language to code: an unsupervised ontology based approach. . In: *IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE-2018)*. Laguna Hills, California, USA, 26-28 September 2018. Laguna Hills, pp.1-8. [online] Avialable at: <https://ieeexplore.ieee.org/xpl/conhome/8526512/proceeding> [Accessed 08 October 2024].

Maulud, D.H., Zeebaree, S.R.M., Jacksi ,K., Sadeeq, M.A.M., Sharf ,K.H., 2021. A State of Art for Semantic Analysis of Natural Language Processing. *Oubahan Academc Journal*, 1(2), pp. 21-28. https://doi.org/10.48161/qaj.v1n2a44 [online] Avialable at: <https://www.researchgate.net/publication/350521473_A_State_of_Art_for_Semantic_Analysis_of_Natural_Language_Processing> [Accessed 08 October 2024].

Tkachenko, K., 2024. Using of NLP Methods in Intelligent Educational System. *Digital platform: Information technologies in sociocultural sphere,* 7(1), pp. 80-96. https://doi.org/10.31866/2617-796X.7.1.2024.307009

Li, W., 2020. Analysis of Semantic Comprehension Algorithms of Natural Language Based on Robot's Questions and Answers. In: *IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA-2020)*. pp. 1021-1024.

Tkachenko, O., Tkachenko, O., Tkachenko, К., 2020. Designing Complex Intelligent Systems on the Basis of Ontological Models. In: *CMIS-2020: Third International Workshop on Computer Modeling and Intelligent Systems.* Zaporozhye, Ukraine, 27 April-1 May 2020. Zaporozhye, pp.266-277.

Semantic Analysis: Definition and Use Cases in Natural Language Processing, n.d. [online] Avialable at: <https://datascientest.com/en/semantic-analysis-definition-and-use-cases-in-natural-language-processing> [Accessed 08 October 2024].

Sowa, J.F., 2002. Concepts in Lexicon: Introduction. Architectures for Intelligent Systems. *IBM Systems Journal*, 41(3), pp. 331-349.

Palagin, A., Gladun, V., Petrenko, N., Velychko, V., Sevruk, A., Mikhailyuk, A., 2008. Informational model of natural language processing. *International Journal Information Technologies & Knowledge*, 2(1), pp. 5-6.

Mochalova, A.V., Mochalov, V.A., 2016. Mathematical model of an ontological-semantic analyzer using basic ontological-semantic patterns. In: Proceedings of 15th Mexican International Conference on Artificial Intelligence, pp. 53-66.

Fu, R. et al., 2014. Learning semantic hierarchies via word embeddings. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 1, pp. 1199-1209.

Eremin, E., 2008. About ontology application to the description of syllabus. *International Journal Information Technologies & Knowledge*, 2(5), pp. 449-454

Gavrilova, T.A., Voinov, A.V., 2007. The cognitive approach to the creation of ontology. *Automatic Documentation and Mathematical Linguistics*, 4 (12), pp. 59-64.

Gladun, V., Velychko, V., Svyatogor, L., 2009. Hierarchical three-level ontology for text processing. *International Journal Information Technologies & Knowledge,* 3(4), pp.303-311.

Lanin, V., 2015. Integration of ontology resources into open format documents for semantic indexing. *International Journal «Information Technologies & Knowledge*, 9(4), pp. 389-397.

Sanfilippo, E.M., 2018. Feature-based product modelling: an ontological approach. *International Journal of Computer Integrated Manufacturing*, 31 (11), pp.1097-1110.

List, C., 2018. Levels: descriptive, explanatory, and ontological. *LSE Research Online*. [online] Avialable at: <http://eprints.lse.ac.uk/87591/1/List_Levels%20descriptive_2018.pdf> [Accessed 07 October 2024].

Gelfert, A., 2017. The Ontology of Models, In: Magnani, L. and Bertolotti, T. eds. *Springer Handbook of Model-Based Science*. Springer, pp.5-23.

Lytvyn, V., Vysotska, V., Dosyn, D., Lozynska, O. and Oborska, O., 2018. Methods of Building Intelligent Decision Support Systems Based on Adaptive Ontology. In: *IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*. Lviv, Ukraine, 21-25 August 2018. Lviv, pp.145-150.

Hermann, K.M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., Blunsom, P., 2015. Teaching machines to read and comprehend. Advances in Neural Information Processing Systems. 28, pp. 1684-1692. [online] Avialable at: <https://proceedings.neurips.cc/paper_files/paper/2015/file/afdec7005cc9f14302cd0474fd0f3c96-Paper.pdf> [Accessed 07 October 2024].

Костянтин Ткаченко,
кандидат економічних наук, доцент,
доцент кафедри програмного забезпечення
комп'ютерних систем,
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»,
Київ, Україна
tkachenko.kostyantyn@gmail.com
https://orcid.org/0000-0003-0549-3396

СЕМАНТИЧНИЙ АНАЛІЗ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ: ОНТОЛОГІЧНИЙ ПІДХІД

Розвиток інформаційних (інтелектуальних) систем навчання, систем електронного документообігу, веб-орієнтованих систем, що працюють з текстовою інформацією природною мовою, призвів до збільшення обсягів навчального контенту та/або масивів оброблених повнотекстових документів. Все це вимагає нових засобів організації доступу до інформації, багато з яких слід віднести до інтелектуальних систем обробки знань. Одним із ефективних підходів до ідентифікації та обробки сенсу навчального контенту (і/або текстових документів) є використання онтологій.

Метою статті є дослідження, аналіз різних підходів до визначення семантичного наповнення текстів природною мовою, розгляд існуючих концепцій аналізу текстів і перспектив використання запропонованого онтологічного підходу до семантичного аналізу текстів природною мовою.

Методами дослідження є методи семантичного аналізу основних понять аналізованої предметної галузі (семантичного анаізу текстів природньої мовою). У статті розглядається підхід до лінгвістичного аналізу текстів, що базується на онтологічному моделюванні.

Новизною дослідження є застосування запропонованого онтологічного підходу до семантичного аналізу текстів природною мовою для визначення сенсу (семантики) текстової інформації, яка використовується в інтелектуальних системах різних класів.

Висновок дослідження, проведеного у статті, полягає в наступному: запропоновано онтологічний підхід до семантичного аналізу тексту природної мови, до його завдань та методів. Використання запропонованого підходу до аналізу текстів приводить до розуміння семантичного аналізу як єдиної тріади: <Онтологія – Текст – Значення>. Для ефективного та коректного вилучення знань у роботі пропонується використовувати багаторівневу онтологію. Результатом взаємодії конкретного тексту природної мови з онтологією є онтологічний сенс – сукупність взаємопов'язаних підграфів онтографа.

Онтологічний зміст витягується з онтографа за допомогою семантичного аналізатора. Діалоговий процесор досліджує синтаксичне дерево розбору речення (зв'язного елемента тексту природної мови) і на основі заданого запитання знаходить фрагмент у тексті, який є відповіддю на запитання. Комп'ютерне розуміння (в інформаційних чи інтелектуальних системах) тексту природної мови досягається, зокрема, за рахунок: занурення тексту в єдине середовище знань – онтологію; формальне представлення сенсу (семантики) в базі знань відповідної системи; можливість операцій над онтологічним змістом.

Запропонований підхід може бути використаний для створення інтелектуальних сховищ інформації, які працюють в єдиному середовищі знань.Запропонований підхід до семантичного, аналізу текстів природною мовою зосереджений на автоматичному вилученні метаданих із текстів різної природи (наприклад, текстового документа, Інтернет-контенту, навчального контенту відповідних онлайн-курсів, опису комп'ютерних та настільних ігор). При подальшому розвитку запропонованого підходу він може використовуватися в системах автоматичного реферування наукових публікацій, змістовної інтерпретації мультимедійного контенту, навчання та тестування (в тому числі з елементами візуального відображення інформації та елементами гейміфікації).

10.10.2024