# Designing Linguistic Ontologies for Training Information Systems

Olha Tkachenko[a], Kostiantyn Tkachenko[a], Oleksandr Tkachenko[b]

[a] *State University of Infrastructure and Technology, I. Ogienko str., 19, Kyiv, 02000, Ukraine*
[b] *National Aviation University, Liubomyra Huzara ave. 1, Kyiv, 03058, Ukraine*

**Abstract**
The article discusses the problems of designing linguistic ontologies for educational information systems. An approach to the formalized description of linguistic ontologies is considered, taking into account the concepts of subject areas of training information systems and the relationship between these concepts. The thesaurus of the training information system, built on the basis of linguistic ontologies, is considered.

**Keywords 1**
training information system, subject area (domain), ontology, linguistic ontology, information resource.

## 1. Introduction

The increase in the volume of text information (electronic documents, web content, educational and methodological material of training information systems, etc.) provides the need for processing such unstructured information, improving the quality and efficiency of existing methods of processing words and developing new ones.

Among the directions of processing unstructured text information, one can single out, for example:

- search for information;
- classification clustering of text documents,
- filtering, rubrication of text documents,
- annotation of a document (group of documents);
- search for similar documents and duplicates,
- document segmentation;
- assessment of semantic similarity and kinship;
  - extraction of information;
  - recognition of named entities;
  - extraction of relationships;
  - extraction of facts;
  - extraction of knowledge;
  - co-reference permission;
- answers to questions in natural language;
- machine translate;
- summary of the text;
- analysis of the sentiment of the test document;
- intellectual analysis;
- automatic creation of ontologies / dictionaries / thesaurus / knowledge base;
- speech recognition and speech synthesis.

COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems, April 22–23, 2021, Kharkiv, Ukraine
EMAIL: oitkachen@gmail.com (O. Tkachenko); tkachenko.kostyantyn@gmail.com (K. Tkachenko); aatokg@gmail.com (O. Tkachenko)
ORCID: 0000-0003-1800-618X (O. Tkachenko); 0000-0003-0549-3396 (K. Tkachenko); 0000-0001-6911-2770 (O. Tkachenko)

Modern educational information systems work with textual information of subject areas, which include thousands of different classes of entities that are among themselves in a huge number of different types of relationships [1, 2].

Therefore, the methods of processing textual information in such systems are often guided by the use of statistical characteristics of this information, in particular:

- frequency of occurrence of words in a sentence, text, set of documents (educational materials, test items, reference information, etc.);
- joint occurrence of words.

Such methods use minimal knowledge about the subject area (domain), language, its features and diversity.

User of training information systems (lecturer, methodologist, student), performing text information processing, primarily:

- reveals the main content of the document and the meaning of its key concepts;
- the main topic, subtopics and key concepts of the document (educational materials, test items, reference information, etc.).

For this, the user of training information systems (lecturer, methodologist, student) usually uses a large amount of knowledge about:

- language of presentation of educational materials, test items, reference information (linguistic knowledge);
- subject area (ontological knowledge);
- organization of coherent text (relations between units of knowledge).

Lack of linguistic and ontological knowledge leads to a variety of problems when, for example:

- ways of formulating queries differ from templates for describing relevant situations in documents that are supported by training information systems;
- long requests are processed (for example, when referring to help information);
- the context of the language (individual words and expressions used in the query) is not fully taken into account.

Thus, modern intellectual training systems for processing text information (or training information systems with elements of intellectualization) face the following problems [3]:

- processing of text information of online courses in the considered subject area;
- taking into account the linguistic features of the language and the structure of the corresponding training or test text.

These problems are especially acute in information retrieval systems, automatic text processing systems (including their generation) and training information systems.

Intellectual text analysis is one of the key tasks in the field of artificial intelligence associated with the problems of automatic analysis and synthesis of natural language arising from the interaction of a user (lecturer, methodologist student) with a training information system.

The solution to these problems is closely related to the use of various approaches of artificial intelligence and computational linguistics.

The development of ontological modeling and machine learning methods has made it possible to achieve the quality required for practical use in natural language processing tasks in training information systems.

The use of additional linguistic and ontological knowledge in the automatic processing of texts in training information systems is a difficult task.

This is due to the fact that such knowledge should be described in specially created resources (thesauri, ontologies), which should contain descriptions of a large number of words and phrases and be able to logically derive new knowledge.

When using such resources, it is usually necessary to solve the problem of word ambiguity, i.e. choose their correct value.

The paper considers the extraction of information from the text, which can be used to create formal models of specific areas of knowledge.

In work, this is the area of training courses in the disciplines "Informatics" and "Information systems and technologies".

A simplified approach to language modeling includes various statistical models based on distribution semantics.

This approach determines the semantic similarity between two linguistic elements (such as words or phrases) based on their distribution properties in large fragments of educational methodological or test text without specific knowledge of the lexical or grammatical meanings of the elements.

One of the ways to represent words with this approach is to cut documents into sets and sequences of words − n-grams [4, 5], which take into account the information contained in verbose constructions of length n (bigrams for word pairs, etc.).

N-gram (for n = 1) ignores all properties of an educational-methodical or test document, except for the number of words in it.

A word set is a collection of documents in the form of a matrix, the rows of which correspond to the documents, and the columns to a specific term.

Intersection values describe the number of words in a particular document.

For n ≥ 1, constructions of several words contain additional information (phrases, idioms, etc.) in comparison with a set of single words.

These models often include a weight for each term-document pair.

The indicator is the number of occurrences (frequency) of a term in each document or the probability of finding a word in a document.

This rates the more general words as more important, although this is not always the case.

It is more common to weigh n-grams so that the weight of a word in a certain document is proportional to its quantity in a given document and at the same time is inversely proportional to the frequency of using this word in other documents from the same collection [5].

One of the paradigms of computer resources for training information systems are formal ontologies (for example, the Semantic Web [6, 7]).

But the automatic processing of unstructured natural language texts is difficult to carry out using formal ontologies [8, 12, 13].

Therefore, for the automatic processing of texts, special ontologies (terminological, lightweight, linguistic) are developed [9, 10, 11], in which concepts are not always strictly formalized.

Linguistic ontology is an ontology, the concepts of which are largely associated with the meanings of linguistic units, terms of the subject area [12, 13, 14, 15].

Linguistic ontologies cover most of the words of a language or subject area and at the same time have an ontological structure that manifests itself in the relationship between concepts.

Therefore, linguistic ontologies can be considered as a special type of lexical database and a special type of ontology.

The paper describes a linguistic ontology designed for automatic text processing for the considered subject area, and the resources that are developed on the basis of this ontology.

## 2. Formalized Linguistic Ontologies

The following can serve as a formal definition of ontologies:

$$O = < C, E, At, R, A >,$$

where:

- $C$ − concepts (classes) of ontology;
- $E$ − instances of ontology;
- $At$ − attributes of concepts and instances of ontology;
- $R$ − relations between concepts;
- $A$ − axioms of ontology.

Formalized ontologies consider various computer resources, in particular, rubricators or thesauri. Typically, rubricators do not include instances and attributes, i.e. the formal model of rubricators is a model of the form:

$$O = < C, R, A >.$$

Formalized ontologies are logical theories built on axioms. To describe them are used:

- logics: descriptive logics, modal logics, first-order predicate logic;
- ontology description languages: DAML + OIL, OWL, CycL, Ontolingua [16, 17, 18].

Ontologies (thesauri, rubricators), the concepts of which are not fully defined in terms of formal properties and axioms, are called lightweight ontologies.

There are different interpretations of the relationship between ontology and the natural language of documents of the training information system:

- ontology is a structure independent of natural language;
- ontology is a structure that is independent of a specific natural language;
- elements of the language lexicon are included in the formal definition of ontology;
- the formal definition of ontology includes the entire lexicon of the subject area (domain).

Based on the foregoing, the formal model of ontology can be described as:

$$O = < V, C, R_{VC}, R_{VR}, R, A >$$

where:

- $V = V_C \cup V_R$ – vocabulary of ontology, containing a set of lexical units for $V_C$ concepts and $V_R$ relations;
- C – a set of concepts of ontology;
- $R_{VC}$ – a set of connections between lexical units $\{v_j\} \subset V$ and the corresponding concepts from $\{c_k\} \subset C$ and relations of the given ontology;
- $R_{VR}$ – the set of links between lexical units $\{v_j\} \subset V$ and the corresponding relations $\{r_i\} \subset R$ of the given ontology;
- $R$ is the set of relationships between concepts of ontology;
- $A$ is a set of ontology axioms.

In the considered formal approaches, words of a natural language are one of the components of the ontological model, lexical expressions are presented only as auxiliary elements that name the concepts and relations of the ontology.

Establishing relationships between concepts, words and expressions of a natural language has many problems, in particular, the introduction of a new concept into an ontology must be associated with existing linguistic elements; definition of relations "concept − linguistic element".

Therefore, a large number of widely known medical ontological resources are thesauri that do not have a high degree of formalization of their structure.

Thesauri are linguistic ontologies, i.e. ontologies based on the meanings of real natural language expressions.

Training information system thesaurus is a normative vocabulary of terms in natural language that explicitly indicates the relationship between terms and is intended to describe the content of documents and search queries.

The basic unit of thesauri is terms, which are categorized into descriptors (= authorized terms) and non-descriptors (= ascriptors).

At their core, descriptors unambiguously correspond to the concepts of the subject area (domain). Relationships between descriptors are divided into: hierarchical and associative.

Hierarchical relationships are usually viewed as asymmetric and transitive.

Hierarchical relationships used in teaching information systems thesauri:

- class − subclass (predecessor − successor, above − below) − is installed between two descriptors, if the concept of a lower − level descriptor (successor, subclass) is included in the concept of a superior descriptor (predecessor, class);
- whole − part.

The purpose of developing training information systems thesauri is to use their units (descriptors) to describe the main topics of documents in the process of manual indexing.

Therefore, it is important that the set of thesaurus descriptors allow describing the topics of educational, methodological, test and reference documents of the subject area.

In this case, the indexing process for such a thesaurus is based on linguistic, grammatical knowledge, as well as knowledge of the subject area.

To determine the semantics of the document test, the component of the training information system – the program “Indexer” – must first read the text, understand it and then state the content of the text using the descriptors specified in the thesaurus.

The program “Indexer” should have a good understanding of all the terminology used in the text −

to describe the main topic of the text, he will need a much smaller number of terms.

The presence of the program “Indexer” testifies to the intellectualization of the training information system.

Thus, the formal model of the thesaurus ($T$) of the training information system can be represented as follows:

$$T = <D, C, R, A>,$$

where:

- $D$ is a set of domain descriptors corresponding to the concepts of a given domain, which are necessary to express the main topics of documents in this domain;
- $C$ − a set of terms (concepts) of the subject area: области: $D \subset C$;
- $R$ – relations of the thesaurus, $R = R_I \cup R_A$ ($R_I$ – hierarchical and $R_A$ – associative relations of the thesaurus);
- $A$ – axioms of transitivity of hierarchical relations.

The described model of the thesaurus of the training information system is intended for its use documents in the process of expert analysis of educational, methodological, test and reference documents.

A thesaurus intended for automatic text processing should contain much more information about the structure and language of the subject area.

The relationships between the terms specified in the thesaurus should be formalized for their use in the training information system.

## 3. Linguistic Ontologies in the Training Information Systems

Formal ontologies (with their independence from a particular language) are difficult to use in automatic text processing for information retrieval applications because:

- units of formal ontology must be associated with units of a specific natural language;
- the desire for a clear formalization of relations between concepts in a formal ontology is difficult to observe when creating super-large resources;
- leads to problems in establishing relations “concept − linguistic expression”.

An training information system deals not only with general vocabulary, but also with specific subject areas and their terminologies.

The description of the terminology of the subject areas of training information systems should use:

- information retrieval context;
- resource units, which are created based on the values of terms;
- description of verbose expressions; principles of inclusion (non-inclusion) of verbose units;
- a small set of relationships between conceptual units.

The use of a linguistic resource in automatic text processing in a training information system should take into account the following provisions:

- conceptual units are created based on the meanings of real linguistic expressions;
- multi-step hierarchical construction of the lexical and terminological system of concepts;
- principles of describing the meanings of polysemous words and expressions;
- development of linguistic ontology as a hierarchical system; the use of formally defined relations with formal properties;
- the use of transitivity and inheritance of relations between concepts of domain as axioms (inference rules).

The LO linguistic ontology model for the SA subject area can be represented as follows:

$$LO = < C, E, N, R, P_{tr}\, T, S, W, L, T_W >$$

where:

- $C$ − a set of concepts of ontology, where concept is a class of entities that have the same properties and relationships with other classes of entities;
- $E$ − a set of instances of ontology concepts, a mapping $E: C \rightarrow E$ is given;
- $N$ − a set of unique names of concepts and instances in the ontology;



COLINS’2021, Volume II: Workshop. Kharkiv, Ukraine, April 22-23, 2021, ISSN 2523-4013
http://colins.in.ua, online

- $R$ is a set of relationships between concepts;
- $P_{tr}$ − set of withdrawal rules;
- $T$ − a set of linguistic expressions, the values of which are presented in the ontology;
- $S$ − a set of relations between linguistic expressions ($T$) and concepts ($C$): $\{s\ (c_i, t_j)\}$;
- $W$ − a set of polysemous words and expressions from $T$: $W \subset T$; $W = W_m \cup W_a$, where $W_m$ are text inputs that refer to more than one concept of the ontology, and $W_a$ are multivalued text inputs that are represented in the ontology by only one value;
- $L$ − a set of lemmatic representations of a linguistic expression (for example, the phrase information system is presented in a lemmatic form as an INFORMATION SYSTEM);
- $T_W$ is a mapping of the terminological composition of a given subject area to text inputs and ontology concepts.

The proposed linguistic ontology of the subject area is a knowledge base of the ontological type about the conceptual system, the lexical and terminological composition of the subject area (disciplines “Informatics” and “Information systems and technologies”), supported by the corresponding training information system.

The unit of linguistic ontology is a concept, as a unit in a system of concepts, which has its own specific properties that distinguish this unit from other units in the system of concepts.

Each entered concept must have a unique name. The name can be an unambiguous word or phrase, the meaning of which corresponds to this concept.

Each concept is supplied with a set of text inputs − language expressions, the values of which correspond to the given concept. Such linguistic expressions are ontological synonyms among themselves.

The texts may contain many variants of text inputs of a particular concept.

The developer of a training information system or a specific online training course must record these options immediately when entering a concept, or supplement it when found in a specific text.

In the texts of the subject area, a significant part is made up of words that belong not only to a specific subject area, but also to the general vocabulary of many subject areas, for example, *create*, *participate*, *accept*, *evaluate*, etc.

Therefore, the polysemantic words described in the linguistic model are divided on:

- the set $W_m$, which includes expressions related to two or more concepts;
- the set $W_a$, which includes expressions related to one concept, but these words may have a different meaning in the general lexicon, which is marked by a special mark of ambiguity.

Relationships between concepts from an ontological resource should perform the following functions:

- these relations should be used in the classic functions of information retrieval thesauri to expand a search query or display a heading of a document;
- relations should be used to resolve the ambiguity of linguistic units included in the resource;
- relations in an ontological resource can be used to identify lexical connectivity in texts and to use the revealed text structure to improve the quality of text processing.

When creating a linguistic ontology of large magnitude, for processing texts that are not limited in style, genre, size, the most stable way is to rely on relationships that do not disappear, do not change during the entire lifetime of any or the vast majority of instances of the concept: for example, software is always consists of programs.

Therefore, in linguistic ontology, relations are described only between such concepts ci and cj, which are inherent in at least one of these concepts by definition.

The properties of transitivity and inheritance are used as axioms.

For a logical conclusion when processing texts in the subject area, it is necessary to describe the relationship between concepts that retain their significance, reliability in various contexts of mentioning concepts.

The main relations in the proposed linguistic ontology are:

- class-subclass;
- whole-part;
- relation of ontological dependence (asymmetric association);

- symmetrical association.

Let the class–subclass ($c_i$, $c_j$) be the relationship between the concepts $c_i$ and $c_j$ ($c_i$ is a subclass of $c_j$), $r(c_i, c_j)$ be an arbitrary relationship between the concepts $c_i$ and $c_j$.

Class–subclass relationships have transitivity and inheritance properties.

However, the same expressions of natural language can correspond to different relationships between entities of the subject area, including those with completely different properties [13. 19].

Therefore, you should check the established class-subclass relationship. For example, to check the belonging of instances of a lower-level concept $c_i$ to a set of instances of a higher-level concept, which implies an answer to the question:

If an object is an instance of one concept, then will it necessarily be an instance of some other concept $c_j$?

The feature of the whole-part relationship is one of the most famous and useful in various subject areas. *Part–whole* relationship is the variety of its manifestations. The most typical objects to which this relation applies are physical objects, entities that last in time, groups of entities, processes, etc.

When modeling this relationship in computer resources, it is important to ensure its transitivity. When describing the whole–part relationship in the proposed model of linguistic ontology, efforts were made to ensure the transitivity of this relationship. That is, it is necessary to describe the whole–part relationship as follows:

*if the text (a fragment of the text) is devoted to the discussion of a part, then it can be assumed that the text (a fragment of the text) will be relevant to the discussion of the whole.*

The condition for ensuring such inheritance is the ontological dependence of the existence of a part on the existence of the whole.

The part dependency can be like this:

- *in existence*, when an instance of a part cannot be separated from an instance–whole;
- *generic*, in which the existence of an instance–part requires the existence of at least one instance of the whole.

The description of hierarchical relationships should be independent of the context in which they are mentioned.

This is important in automatic text processing, since in automatic mode it is often impossible to use the context to confirm the existence of a particular relationship.

In linguistic ontologies, the following properties of the whole-part relationship are used:

- part $(c_1, c_2) \leftrightarrow$ whole $(c_2, c_1)$;
- whole $(c_1, c_2) \wedge$ whole $(c_2, c_3) \rightarrow$ whole $(c_1, c_3)$ – transitivity of the relation;
- class $(c_1, c_2) \wedge$ whole $(c_2, c_3) \rightarrow$ whole $(c_1, c_3)$ – inheritance of the whole relation with respect to the class–subclass relation.

The concept $c_i$ is externally dependent on the concept $c_j$ if for all instances of $c_i$ there is an instance $c_j$ that is not part or material of the instance $c_i$.

For example, the concept of a son is externally dependent on the concept of a parent, since it exists only within the family in relation to its parents.

And the concept of a car is not externally dependent on any entity, since it requires the existence of a motor, which is part of the car.

The asymmetric association relation Ass represents an external ontological relationship between concepts. This relationship is established between the concepts $c_1$ and $c_2$ if the following conditions are satisfied:

- between the concepts $c_1$ and $c_2$, the class-subclass and / or whole-part relations cannot be established;
- the statement is true: the existence of $c_2$ means the existence of $c_1$.

These conditions mean that the dependent concept $c_2$ is externally dependent on $c_1$:

$$Ass_1(c_2, c_1) = Ass_2(c_1, c_2).$$

Ontological dependency relationships are applicable to different areas, so they are most often used in top-level ontologies.

For various applications of automatic word processing, some groupings of concepts and relations in linguistic ontology are used.

## 4. Linguistic Ontologies Based on the Described Model

The above principles were the basis for the development of an ontology for the disciplines "Information systems and technologies" and "Informatics".

The created ontological resources have the same structure. They are ontologies because they describe the concepts of the domain and the relationship between them.

These resources belong to linguistic ontologies, since the introduction of concepts is largely motivated by the meanings of linguistic units related to the subject area of the resource.

At the same time, they are thesauri, since each concept is associated with a set of linguistic expressions (words, terms, phrases) with which this concept can be expressed in a text - such a set of textual concept inputs is necessary to use ontologies for automatic text processing.

Each term is provided with a description (dictionary entry), has hierarchical links with other terms and synonyms.

Figure 1 shows a list of hyperlinks to dictionary entries of the "main root" key terms (concepts) of the subject area "Information systems and technologies" and "Informatics".

Having opened the dictionary entry of a term, we get a description of the term, a list of other related terms and lists of publications and persons related to this term. The performed layout allows you to view the thesaurus in alphabetical order of its text inputs.

The choice of a specific text input, for example, TECHNOLOGY, allows you to see the totality of concepts to which this word is attributed, namely to the concepts of INFORMATION TECHNOLOGY and INFORMATION SYSTEM.

| | |
|---|---|
| Алгоритм | Algorithm |
| Архітектура обчислювальної машини | Computer architecture |
| База даних | Database |
| Дані | Data |
| Знання | Knowledge |
| Інформатика | Informatics |
| Інформаційна система (ІС) | Information system (IS) |
| Інформаційна технологія (ІТ) | Information Technology (IT) |
| Інформаційні ресурси | Information resources |
| Інформаційний пошук | Information search |
| Інформація | Information |
| ІТ-бізнес | IT business |
| Кібернетика | Cybernetics |
| Комп'ютер | Computer |
| Комп'ютерна мережа | Computer network |
| Мова програмування | Programming language |
| Модель | Model |
| Обчислювальна система | Computing system |
| Операційна система (ОС) | Operating system (OS) |
| Програмний код | Program code |
| Програмне забезпечення | Software |
| Програмування | Programming |
| Система | System |
| Технологія | Technology |

**Figure 1**: Key Concepts of Disciplines "Informatics" and "Information Systems and Technologies"

For each concept, complete lists of text inputs are indicated, including words of different parts of speech, as well as phrases. So, for the concept INFORMATION TECHNOLOGY, the text inputs are words and expressions: *technology*, *information*, *software*, *information resources*, *information system* (Figure 2).

| | | |
|---|---|---|
| WHOLE | Інформаційна технологія | Information technology |
| PART | Інформатизація | Informatization |
| ASSOCIATION | Інформаційна система | Information system |
| PART | Цифровізація | Digitization |
| ASSOCIATION | Інформаційна революція | Information revolution |
| PART | Інформаційні ресурси | Information resources |
| ASSOCIATION | Мова програмування | Programming language |
| PART | Програмування | Programming |
| PART | Програмний код | Program code |
| PART | Програма | Program |
| ASSOCIATION | Обчислювальна система | Computing system |
| ASSOCIATION | Обчислювальна техніка | Computers |

**Figure 2**: Article of the Concept INFORMATION TECHNOLOGY

For each concept, relationships with other concepts are indicated. In Figure 2, in the article the concepts of INFORMATION TECHNOLOGY are indicated:

- parts of the concept INFORMATION TECHNOLOGY (INFORMATIZATION, DIGITIZATION, INFORMATION RESOURCES, PROGRAMMING, PROGRAM CODE, PROGRAM, etc.);
- ontologically dependent concepts, i.e. concepts that could not have appeared if information technology did not exist: INFORMATION SYSTEM, INFORMATION REVOLUTION, PROGRAMMING LANGUAGE, COMPUTING SYSTEM, COMPUTERS, etc.

Figure 3 shows the concepts related to the key concept of ALGORITHM and which are in different types of relationships with it.

| | | |
|---|---|---|
| CLASS | Алгоритм | Algorithm |
| SUBCLASS | Алгоритм Дейкстри | Dijkstry's algorithm |
| PART | Блок-схема | Block diagram |
| SUBCLASS | Машина Т'юринга | Turing machine |
| SUBCLASS | Машина Поста | Post Machine |
| ASSOCIATION | Мова програмування | Programming language |
| SUBCLASS | Нормальній алгорифм Маркова | Normal Markov algorithm |
| PART | Програмний код | Program code |
| ASSOCIATION | Програма | Program |
| SUBCLASS | Примітивно-рекурсивна функція | Primitive-recursive function |
| PART | Складність алгоритму | Complexity of the algorithm |

**Figure 3**: Article of the Concept ALGORITHM

For each concept, relationships with other concepts are indicated. In Figure 3 in the article of the concepts of ALGORITHM are indicated:

- types of algorithms formalization (DIJKSTRY'S ALGORITHM, TURING MACHINE POST MACHINE NORMAL MARKOV ALGORITHM PRIMITIVE-RECURSIVE FUNCTION);
- parts of the concept of ALGORITHM (BLOCK DIAGRAM, COMPLEXITY OF THE ALGORITHM, etc.);

- ontologically dependent concepts, i.e. concepts that could not have appeared if there were no algorithms: PROGRAM, PROGRAMMING, PROGRAM CODE, PROGRAMMING LANGUAGE, etc.

The information base supporting the proposed linguistic ontology includes:

- set of concepts for the subject area under consideration (disciplines "Informatics" and "Information systems and technologies", which are supported by the corresponding training information system):
    - concepts of general vocabulary;
    - concepts of subject areas "Informatics" and "Information systems and technologies";
- interpretation of concepts;
- set of relationships between the concepts of the considered subject area;
- many text inputs of the thesaurus; • description of text inputs:
    - lemmatical representation of text input;
    - syntactic type;
    - the main word of the noun phrase;
- set of correspondences of text inputs to the concepts of the thesaurus of the training information system.

## 5. Conclusion

The article presents a model of linguistic ontology for the subject area (disciplines “Informatics” and “Information systems and technologies”).

This model is used in the development of a training information system that supports online learning in these disciplines.

In the proposed model, a set of relations of a linguistic ontology is described, which is specially selected to describe the subject area under consideration.

The functions of relations of the linguistic ontology of information retrieval are possible when providing multi-step logical inference based on the properties of transitivity and inheritance of relations and their independence from the context of the concept.

To provide these properties, it was proposed to use a small set of relations.

Ontological definitions of the relations used were introduced. Such system of relations reflects the most essential relationships between entities and can be used to describe relationships between concepts in a variety of disciplines, supported by educational information systems.

The proposed linguistic ontological model was implemented in the implementation of a training information system that supports the disciplines "Informatics" and "Information systems and technologies."

## 6. References

[1] G.Z. Liu, A Key Step to Understanding Paradigm Shifts in E-learning: Towards Context-Aware Ubiquitous Learning, British Journal of Educational Technology, 2017. Vol. 41. № 2. pp. E1-E9.

[2] V.A. Trainev, New information communication technologies in education, Moscow, Dashkov and K, 2018. (in Russian).

[3] Matthew U. Scherer, Regulating artificial intelligence systems: risks, challenges, competencies, and strategies, Harvard Journal of Law & Technology Vol.29, № 2, Spring 2016.

[4] ISO 25964-1:2011, Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. Geneva: International Organization for Standards, 2011.

[5] G. Miller, Nouns in WordNet. WordNet – An Electronic Lexical Database. The MIT Press, 1998, pp. 23-47.

[6] O. Tkachenko, A. Tkachenko, K.Tkachenko, Ontological Modeling of Situational Management, Digital platform: information technology in the sociocultural area. Vol. 3. № 1 (2020). pp. 22-32.

[7] Kamran Munira, M. Sheraz Anjumb, The use of ontologies for effective knowledge modelling and information retrieval. 2017. URL: https://www.sciencedirect.com/science/article/pii/S2210832717300649. doi: 10.1016/j.aci.2017.07.003.
[8] S. Nirenburg, Y. Wilks, What’s in a symbol: Ontology, representation, and language, Journal of Experimental and Theoretical Artificial Intelligence, 2001. Vol. 13(1). pp. 9-23.
[9] S.E. Greger, S.V. Porshnev, Building an ontology of information system architecture, Fundamental Research, № 10, 2013, pp. 2405-2409.
[10] J. Sowa, Building, Sharing and Merging Ontologies. URL: http://www.jfsowa.com/ontology/ontoshar.htm.
[11] C. List, Levels: descriptive, explanatory, and ontological, 2018. URL: http://eprints.lse.ac.uk/87591/1/List_Levels%20descriptive_2018.pdf.
[12] S. Nirenburg, V. Raskin, Ontological Semantics. MIT Press, 2004. 420 p.
[13] Kudashev, Quality Assurance in Terminology Management: Recommendations from the TermFactory Project. – Helsinki: Unigrafia, 2013. 216 p. URL: http://www.projectglossary.eu/download/QA_in_TM_Kudashev.pdf.
[14] B. Magnini, M. Speranza, Merging Global and Specialized Linguistic Ontologies, Proceedings of OntoLex. (2002). pp. 43-48.
[15] T. Veale, Y. Hao, A context-sensitive framework for lexical ontologies, Knowledge Engineering Review, 2007. Vol. 23(1). pp. 101-115.
[16] I. Horrocks, Reviewing the design of DAML+OIL: An ontology language for the Semantic Web. URL: https://www.researchgate.net/publication/2477217_Reviewing_the_design_of_DAMLOIL_An_ontology_language_for_the_Semantic_Web
[17] Web Ontology Language (OWL). URL: https://www.w3.org/OWL/
[18] Ontolingua. URL: http://www.ksl.stanford.edu/software/ontolingua/
[19] N. Guarino, Some Ontological Principles for Designing Upper Level Lexical Resources, Proceedings of First International Conference on Language Resources and Evaluation. Granada, Spain, 1998.