

Automatic Methods for Extracting Taxonomic Relationships from Texts

N. V. Loukachevitch^{a,*}

^a *Lomonosov Moscow State University, Moscow, 119991 Russian Federation*

**e-mail: louk_nat@mail.ru*

Abstract—An overview of approaches to automatic extraction of taxonomic relationships (is_a relationships) from texts, including classical methods based on lexico-semantic patterns and vector representations of words and their modern development, is provided. It also describes new approaches that are based on language models like BERT and use new options for working with lexico-semantic patterns, in which the model must fill the desired position in the pattern. Various ways to test the quality of extraction of taxonomic relationships and the results of these tests are presented. According to the results currently achieved, it can be concluded that, despite the progress in the development of methods for extracting taxonomic relationships, the quality of extraction is insufficient for fully automatic construction or enrichment of taxonomies.

Keywords: taxonomies, class–subclass relations, hyponym–hypernym, knowledge extraction from texts

DOI: 10.1134/S1054661823030276

INTRODUCTION

Texts in natural language contain large amounts of knowledge about the world, about various subject areas, and about the use of various words and expressions in the language, which enables automatic or automated construction of ontologies from texts [1, 16].

Taxonomic relationships are the most well-known type of relationships that are in demand in the description of many subject areas. These relationships can have different names in different fields of knowledge, such as class–subclass relationships (ontology development), subsumption relationships (information retrieval thesauri), hyponym–hypernym relationships (lexical semantics, WordNet thesaurus [26]), and is_a relationships (artificial intelligence).

Taxonomic relationships form the structural core of the description of a subject area. Therefore, an important task is to automate the construction of taxonomies, as well as make it possible to assign this relationship automatically to new entities that appear in the texts of the subject area and to perform the so-called enrichment of taxonomies (ontologies) based on the texts of the subject area.

This review will consider approaches to the extraction of taxonomic relationships from texts, existing datasets, and the approaches to assessing the quality of relationship extraction, as well as the results achieved thus far.

1. METHODS USED TO EXTRACT TAXONOMIC RELATIONSHIPS FROM TEXTS

The automatic construction of taxonomies from texts has attracted close attention from researchers for almost 30 years [18]. When constructing semantic relationships based on text corpora, two main approaches are distinguished: a linguistic approach based on lexico-semantic patterns and an approach based on distributive semantics. Currently, the combined approaches have also received great development. In addition, in connection with the advent of BERT-type neural network models [13] pre-trained on large corpora, approaches based on automatic filling of patterns of the desired type are being developed. Further in the text, we will use the term hypernym to designate a word with a broad meaning, and the term hyponym for a subordinate word with a narrower meaning.

1.1. Methods Based on Linguistic Approaches

The linguistic approach is the extraction of relationships between words based on their joint occurrences in a certain set of patterns. Methods that use patterns are based on the fact that if some given context is found between entities X and Y, then there is a given relationship between these entities. For a taxonomic relationship, this context may be the context “X is a type of Y.” This approach is highly accurate and simple and can be applied to various types of relation-

ships. Work on extracting taxonomic relationships by patterns began with the study [18], where several of the most frequent patterns were proposed.

The use of patterns for extracting relationships is usually highly accurate, but also involves several problems that reduce the completeness of extracting relationships. First, there may not be a single explicit mention of the contexts of the required type in the text collection used, despite the fact that X and Y are in the required relationship to each other. Because of this, the approach has a very low recall. Second, to achieve a high-quality extraction of relationships, a large number of patterns must be specified. To increase the completeness of lexico-semantic patterns, several methods have been proposed. A number of works have proposed to use an extended set of lexical-semantic patterns. For example, the authors of article [37] use a set of 59 patterns collected from different works.

In order to improve the quality of prediction for low-frequency words, study [31] creates a matrix of mutual occurrence of words in given lexico-semantic patterns, which is estimated using the mutual information formula. The resulting matrix is then subjected to SVD decomposition and dimensionality reduction, the so-called method of latent semantic analysis [12]. As a result, the same hypernyms are predicted for words that appear in similar patterns with a similar set of words, which leads to improvement in the quality of hypernym prediction for low-frequency words or words rarely mentioned in the patterns used.

Another approach to increase the completeness of pattern coverage is the so-called bootstrapping (partial supervised learning), i.e., several patterns for extracting a given relationship are specified manually, and relationship examples extracted based on these patterns are used to collect new types of patterns [41]. However, as the number of patterns increases, the accuracy of relationship extraction usually decreases.

When creating patterns, one may note that some of them have a similar look. In particular, the Patty system [27] increases the completeness of extracting relationships by patterns by generalizing the patterns. For example, the patterns “X inc. is a Y,” “X group is a Y,” “X organization is a Y” can be generalized to “X * is a Y,” which, however, may lead to some decrease in the accuracy of extracting the relationship.

Finally, excessive words can be inserted into lexico-semantic patterns; therefore, a number of works propose to use syntactic sentence structures to extract patterns and write the pattern itself in the language of syntactic relationships [2, 39], which can also connect remotely located words of a syntactic pattern. For the Russian language, lexico-semantic patterns for extracting subsumption relationships and hyponym–hypernym relationships are considered in [8, 20]. The authors of [33] used lexico-semantic patterns to extract hypernyms of words from their interpretations in explanatory dictionaries.

1.2. Methods Based on Vector Representations of Words

The second main direction of extraction of taxonomic relationships is based on the methods of distributive semantics and on the so-called distributive hypothesis, which consists in the fact that the semantic similarity between words (or other linguistic units) can be modeled on the basis of comparison (similarity) of the contexts of these words [17, 22]; i.e., it is assumed that the similarity of word contexts correlates with the semantic similarity of words.

The classic approach of distributive semantics calculates a matrix for each target word w , where each row corresponds to the target word and each column represents the words that occurred in the context of the source word. Thus, each word is represented by a vector of words that occur in its context; i.e., a word is represented geometrically by a point in an n -dimensional space, the dimension of which is equal to the number of different words found in the text collection [14, 43].

The task of finding similarities between words is reduced in this case to determining the proximity of the corresponding points. The proximity of the points can be calculated based on the calculation of the Euclidean distance between the points. However, in this case, the distance will depend strongly on the frequency of the word in the corpus. Therefore, it is assumed that a small angle between vectors emanating from the origin is more important than proximity in the Euclidean distance. To calculate proximity by an angle, one can consider a cosine measure of proximity, in which it is assumed that the smaller the angle, the closer to zero the cosine of this angle. In this case, the similarity between words can be sought using the scalar product between context vectors [14, 43]. The closer the scalar product to 1, the higher the semantic similarity between words.

At present, the approach to constructing vector representations of words based on neural networks has become very popular. It is believed that in many tasks neural network approaches are superior to traditional distributive semantics approaches; in addition, more efficient tools for calculating these representations have been created. In 2013, study [25] proposed an efficient implementation of a method for creating vector representations of words with a relatively small number of dimensions (usually from 100 to 1000) based on neural networks—the word2vec package. The main idea of this approach is to train such vector representations (called word embeddings) on a text corpus in the process of word prediction in context for one or more words. At present, already counted vector representations on large corpora that allow using them in various tasks of automatic text processing without additional computational costs have been published [5, 21].

When extracting taxonomic relationships based on vector representations, it is assumed that hyponyms and hypernyms are used in similar contexts as words that are close in meaning, which must lead to the formation of similar vector representations. However, it should be noted that vector representations of words themselves do not allow one to determine the type of relationship that connects two words. Both synonyms, hyponyms or hypernyms of a target word and antonyms (expensive—cheap) as well as thematically related words (for example, dog—collar), etc., can turn out to be close in vector representations to the original word. Therefore, the original vector representations of words are used as the basis to apply the two basic approaches for extracting hypernyms further: the unsupervised approach and the supervised approach.

In the first group of unsupervised methods, it is assumed that there is some transformation of the original word vectors that better matches the hyponym—hypernym relationship. In particular, it can be assumed that the contexts of the hypernym word include the contexts of the hyponym word—the so-called Distribution Inclusion Hypothesis [32].

The second group of methods (supervised) uses the original vector representations as features for the subsequent application of specialized classifiers trained on hyponym—hypernym datasets [4]. The first supervised approaches used arithmetic operations on word vectors (addition, subtraction, concatenation), to which a classifier (logistic regression, support vector machine, etc.) was then applied to predict a hypernym for a word [4, 32, 34].

As part of the further development of the supervised methods based on the vector space of embeddings, it was supposed that one can try to find a linear transformation that converts a hyponym into a hypernym; this approach is called projection learning [15]. However, further experiments showed that it was impossible to find a single transformation of a hyponym into a hypernym, since the optimal transformation varies for different semantic classes of words. Therefore, methods were proposed for the simultaneous semantic clustering of words and the search for a suitable hyponym—hypernym transformation corresponding to each semantic class [6, 44, 47].

An important role for the above operations in the vector space of words is played by the original corpus, in particular its size and genre of texts (Internet pages, Wikipedia, domain texts, etc.). To improve the quality of vector representations, it was proposed to use an approach based on the so-called meta-embeddings, i.e., unions of various vector representations obtained based on different text corpora [7, 42]. Combinations of vectors such as concatenation or averaging can act as meta-embeddings. A more flexible method is the approach based on the so-called auto-encoders, the task of which, starting from several different vector representations, is to find the best combined represen-

tation (meta-embedding), from which the original representations can be best restored and also to achieve the best correspondence of meta-embeddings to some additional criterion. In the task of predicting hypernyms, such a criterion can be the use of the so-called triplet loss—meta-embedding must be similar to vectors of words that are close in meaning and not similar to vectors of random words [42].

1.3. Combined Methods for Extracting Taxonomic Relationships

Linguistic methods for extracting relationships can use lexico-semantic patterns or syntactic relationships of words [40], the joint occurrence of words in a sentence [48], etc. Methods for extracting relationships based on distributive approaches are based on extracting the contexts of words usage and calculating their similarity as vectors; i.e., these approaches allow one to extract relationships that are not explicitly indicated in the texts.

The advantages of both methods combine the methods of incremental enrichment of taxonomy [40, 47], in which all of the above characteristics (occurrence in patterns, the presence of similarity in different contexts (linear and syntactic, local and global), joint occurrence) serve as signs, on the basis of which a decision is made to add each next word to the created taxonomy. In making decisions about joining a taxonomy, study [40] uses the principle of maximizing the conditional probability of a relationship based on the available features, and the study [48] uses the principle of optimizing the taxonomic structure and modeling the level of abstractness of a word in the taxonomic structure.

The HypeNET [35] model uses three vectors to represent a sentence: two source word vectors and a third vector that is a vector representation for a syntax path (=path in the syntax tree) between two words. The LSTM recurrent neural network is used to create a syntactic path vector. The results of this work significantly surpassed previous results in the task of extracting hypernyms as a classification task in comparison with linguistic and distributive approaches, since the model is also trained to encode syntactic paths between words optimally instead of following predefined lexico-semantic patterns.

1.4. Prediction of Hypernyms by Generating Answers to a Question

In connection with the advent of large pretrained BERT neural network models [13], there has been an improvement in solving many problems of automatic text processing. In this model, it was proposed to use a transformer architecture neural network with a self-attention mechanism [45] for pretraining based on large volumes of unlabeled text data to solve two problems: (1) predicting a masked word in context, (2) pre-

dicting whether the second sentence follows from the first one. As a result, the BERT model creates contextualized (context-aware) representations of words that allow better understanding of the context and solving many problems of automatic text processing.

A specific approach to training the BERT model has given additional opportunities for solving problems of automatic text processing based on a specially asked question (prompt) [34]. This approach has also been used in the problem of predicting hypernyms [46]. In particular, the TaxoPrompt [46] method uses the following pattern as a prompt: “What is parent-of [X]? It is [MASK].”

A new word, for which a hypernym needs to be defined, is inserted into slot X. Further, this pattern is fed into the input of the BERT language model, the learning feature of which is that the model is trained to predict the word hidden by the mask [13]. The following can also be used as alternative prompts: (1) [X], [MASK]; (2) [X] is a [MASK]; (3) [MASK], such as [X].

To give additional context, the described lexical-semantic pattern is fed to the input of the language model together with the second sentence separated by a service token [SEP]. The second sentence could be the interpretation of the target word from a dictionary or the paths of their underlying ontology written as text sequences. To train the model, a training collection with correct and incorrect answers from some taxonomy is built. Thus, in predicting a hypernym, the model uses a huge amount of data on which pretraining took place, interpretations from a dictionary, and a vector representation of a taxonomy.

2. TESTING METHODS FOR EXTRACTING TAXONOMIC RELATIONSHIPS

The quality of extracting taxonomic relationships is assessed using various approaches, which can be divided into four main groups.

(I) Prediction of the relationship of hypernymy as a classification problem on a given dataset. With this formulation, testing of approaches is carried out on the basis of a special dataset. The set contains positive examples between which there is a desired relationship and negative examples between which there is no relationship. Positive examples are usually taken from some well-known ontological resources, and negative examples are generated as a result of some procedure. Thus, the task of testing is posed as the task of classifying a pair of words from a dataset into two classes: whether there is a necessary relationship between words or not. Quality assessment in this task is performed by traditional measures for classification: accuracy, recall, F-measure.

Study [36] presents an overview of various operations on word vectors that were used in hypernym identification problems and testing of hypernym pre-

diction based on several datasets in unsupervised and supervised learning modes. It is shown that unsupervised methods for predicting hypernyms based on several datasets exceed the quality of 0.85, and supervised methods based on some data get ideally correct answers.

Based on the results from such testing, it can be assumed that the problem of determining a hypernym for a word has been solved. However, it has been “solved” only within the framework of the created datasets. The problem with this approach to testing hypernym extraction is that this setting is not practical. In such datasets, the number of positive and negative examples is usually comparable to each other, but in practice the number of taxonomic relationships between words is orders of magnitude smaller than the potential relationships between different pairs of words in a text corpus. For example, one of the well-known and large sets used to test the extraction of hypernym–hypernym relationships is the BLESS dataset [4]. BLESS contains hypernyms for 200 specific, mostly single-valued nouns. Negative pairs contain co-hyponyms, meronyms (words that are in a part–whole relationship), and random pairs of words; as a result, the dataset contains a total of 14542 pairs of words with 1337 positive examples. Hence, there are only ten times fewer positive pairs than negative ones, which does not correspond to the situation with extraction from real text collections.

In addition, a number of studies have shown that commonly used relationship classification methods are not trained to classify a pair of words, but rather to remember the typical role of a word as a hyponym or hypernym. For example, there are words such as animal, plant, or building, which are typical hypernyms and can be used repeatedly as such in datasets [23, 35]. Thus, the results of the methods can be overestimated—this is the so-called problem of lexical memorization [23].

(II) Extracting hypernyms from a given corpus. With such a formulation of testing, the systems must correctly extract hypernyms from the corpus for a given set of words; i.e., the systems must find suitable candidates for hypernyms themselves. In the SemEval-2018 testing [11], this task is posed as a ranking task; i.e., the automatic system must create an ordered list of candidates for hypernyms for the target words. Moreover, the correct candidates must be at the top of the list. The quality of approaches in this problem is assessed using measures such as MRR, the mean reciprocal of the first correct answer, as well as MAP (the mean average precision), which is a combined measure of accuracy and recall and is calculated as the sum of the accuracies at the time of entering the correct answer list, which is averaged over the number of correct answers.

Hypernym prediction methods based on distributive representations were used as the basic methods for comparison in SemEval-2018 testing; when tested on limited datasets, they demonstrated an accuracy above 0.7. In this formulation of the problem, these methods produced the first correct hypernym of a word at an average of 30 positions (MRR 0.03). This is due to the fact that the words that are most similar to the target word according to the vector model often turn out to be not hypernyms, but so-called co-hyponyms (i.e., hyponyms of the same hypernym).

The best result in SemEval-2018 testing was achieved by the CRIM system [6], which used a combined approach that included lexico-semantic patterns for determining hypernyms and co-hyponyms, a projection-learning method with training to transform space from a hyponym to a hypernym with simultaneous training of semantic clusters. This system has achieved a quality of 0.33 MRR, which means that the correct hypernym is predicted at an average of three positions.

In a later approach [3], it is proposed to use not one transition in the hyponym–hypernym matrix transformation as in the original projection learning approach, but also to use such a transformation several times in order to reveal higher hypernyms as in the chain: *Bill Clinton—politician—person*. This approach achieves MRR-0.39 based on the SemEval-2018 data.

(III) Automatic taxonomy generation. The third way of testing is the task of automatically generating a taxonomy based on a text corpus and comparing the generated taxonomy with the existing one. Such a task was posed at the seminars SemEval-2015 and SemEval-2016 [9, 10]. Thus, it is necessary not only to extract the hypernymy relationships for a certain set of target words, but also to build the correct structure of the acyclic graph from them. Existing taxonomies from different areas were used as test data: chemicals, equipment, food, or sciences, ranging from 450 vertices (sciences) to 17000 vertices (chemicals). The vertices themselves are given; it is necessary to determine taxonomic relationships for them. Words can have more than one hypernym, so a taxonomy can be a poly-hierarchy.

This evaluation was conducted as a standard classification task using measures of precision, recall, and F-measure compared to reference resources. Structural quality assessments of the resulting taxonomy (e.g., the absence of cycles) were also conducted. Most participants used Wikipedia as a corpus and applied combined methods, including lexico-semantic patterns and distributive similarity of terms across the corpus. The maximum result of the best system reached an F-measure of 0.39 (an accuracy of 0.36).

Manually checking the accuracy of a taxonomy fragment gave the best result around 0.6.

(IV) Automatic enrichment of taxonomy. With this formulation, it is assumed that there is some general taxonomy and that it needs to be completed in a certain subject area or based on new data. For the first time, a solution to such a problem was proposed and tested on the basis of the completion of WordNet in [40].

In addition, for example, such a task was posed in testing SemEval-2016 task 14 [19]. Test participants had to find the place of a new word in the WordNet hierarchy [26] as a synonym or hyponym for an existing concept. The peculiarity of the task was to complete the WordNet taxonomy at the expense of existing dictionaries, so new words were provided with interpretations from these dictionaries. Quality was measured by two measures: Accuracy, the accuracy of establishing a relationship, taking into account the distance (in the number of relationships) from the correct answer, and Recall, the proportion of words for which the system offered a solution. The results of the best system were slightly better than the results of the basic solution, which chose as the hypernym of the new word the most frequent meaning of the first word of the same part of speech as the new word in the interpretation. Meanwhile, the accuracy of the basic method was a little more than 0.5; answers were given for all words, which led to the value of the F-measure of 0.68.

The current best results in testing SemEval-2016 task 14 were obtained by the TaxoExpan system [38], which achieved an accuracy of 0.566 and an F-measure of 0.723. The system represents the new word and candidate words in WordNet as fastText embeddings, the interpretation is represented as an average vector of embeddings of the mentioned words. The representation of a candidate word as the average of the token and interpretation vectors is fed into a suitable hypernym selection system based on the representation of concepts in the WordNet taxonomy using graph neural networks (GNNs).

In 2020, as part of the Dialogue conference, open testing was organized to enrich the RUSSE-2020 taxonomy for the Russian language [28]. Testing used two versions of the Russian version of WordNet—RuWordNet [24]. The dataset for enrichment consisted of words from an unpublished new version of RuWordNet (RuWordNet 2.0), and the participating systems had to choose a suitable hypernym (more precisely, a synset, a concept in RuWordNet) in the published version of RuWordNet. Both the hypernyms specified in RuWordNet 2.0 and their hypernyms that were higher by one step (second-order hypernyms) were counted as correct answers, thus modeling the work of the “intelligent” expert assistant, within which an approximate place in the semantic network must be found to bind a new word. The prediction quality was

assessed using measures of the average accuracy and MRR.

The basic component used in the work of most participants was the extraction of the most similar words from RuWordNet for the target word based on vector representations; synsets, hypernyms, and second-order hypernyms were used to form candidate synsets as hypernyms for the target word. The same approach was used to create the basic solution provided by the organizers. Many participants noted that the most similar words for fairly specific words that were included in the dataset for testing were not synonyms or hypernyms, but co-hyponyms, i.e., hyponyms of the same hypernym.

Fasttext embeddings [5], which were calculated on the Internet corpus of Common Crawl and Wikipedia,¹ turned out to be the best vector representation for predicting hypernyms. The best method in this testing was the method that uses a variety of features, including candidates based on embeddings, dictionary interpretations of Wiktionary, and analysis of search results in the Yandex and Google search engines. This approach achieved a result of 0.55 MRR for noun hypernym prediction, which, on average, corresponds to finding the first correct answer in the second place of search results.

Continuing this testing and following its main idea (testing taxonomy enrichment based on different versions of the same resource), a specialized dataset was created in 2021 using three consecutive versions of the English WordNet and two versions of RuWordNet—the dataset was called Diachronic wordnets [29]. The datasets include new words from later versions that need to predict the synset–hypernym using an older version of the resources. Thus, based on this dataset, it is possible to test different approaches to enriching WordNet resources in parallel in two languages.

Using the created datasets, the best results were obtained based on meta-embeddings that combine different vector representations, including word embeddings obtained by different methods based on different corpora, vector representations of vertices of enriched resources, and information from Wiktionary interpretations. In particular, the best result on the RUSSE-2020 dataset was increased to 0.6 MRR without any use of the results of Internet search engines [29].

CONCLUSIONS

The taxonomic relationship that links an entity to its class is a central relationship in the representation of knowledge in many subject areas. Therefore, great attention is paid to automatic methods for identifying taxonomic relationships from texts.

¹ <https://fasttext.cc/docs/en/crawl-vectors.html>.

In this article, we have presented an overview of approaches to automatic extraction of taxonomic relationships between words, starting from the most classic ones, such as lexico-semantic patterns. Currently, the basis for the extraction of taxonomic relationships are vector representations of words, which reflect the similarity of the contexts of mentioning a hyponym and its hypernym. This article also presents new approaches that are based on BERT language models and use new options for working with lexico-semantic patterns, in which the model fills the position of a hypernym.

This article also describes various methods for testing the quality of extracting taxonomic relationships and presents the results of these tests, from which it follows that, despite the progress in the development of methods for extracting taxonomic relationships, the quality of extraction is insufficient for fully automatic construction or completion of taxonomies.

CONFLICT OF INTEREST

The author declares that she has no conflicts of interest.

REFERENCES

1. F. N. Al-Aswadi, H. Y. Chan, and K. H. Gan, “Automatic ontology construction from text: A review from shallow to deep learning trend,” *Artif. Intell. Rev.* **53**, 3901–3928 (2020). <https://doi.org/10.1007/s10462-019-09782-9>
2. A. I. Aldine, M. Harzallah, B. Giuseppe, N. Béchet, and A. Faour, “Redefining Hearst patterns by using dependency relations,” in *Proc. 10th Int. Joint Conf. on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Knowledge Engineering and Knowledge Management, Vol. 2 (SCITEPRESS - Science and Technology Publications, 2018), pp. 148–155. <https://doi.org/10.5220/0006962201480155>
3. Yu. Bai, R. Zhang, F. Kong, J. Chen, and Yo. Mao, “Hypernym discovery via a recurrent mapping model,” in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (Association for Computational Linguistics, 2021), pp. 2912–2921. <https://doi.org/10.18653/v1/2021.findings-acl.257>
4. M. Baroni, R. Bernardi, N. Do, and Ch. Shan, “Entailment above the word level in distributional semantics,” in *Proc. 13th Conf. of the European Chapter of the Association for Computational Linguistics, Avignon, France, 2012* (Association for Computational Linguistics, 2012), pp. 23–32. <https://aclanthology.org/E12-1004>.
5. P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” *Trans. Assoc. Comput. Linguist.* **5**, 135–146 (2017). https://doi.org/10.1162/tacl_a_00051

6. G. Bernier-Colborne and C. Barrière, "CRIM at SemEval-2018 Task 9: A hybrid approach to hypernym discovery," in *Proc. 12th Int. Workshop on Semantic Evaluation, New Orleans, 2018* (Association for Computational Linguistics, 2018), pp. 725–731. <https://doi.org/10.18653/v1/s18-1116>
7. D. Bollegala and C. Bao, "Learning word meta-embeddings by autoencoding," in *Proc. 27th Int. Conf. on Computational Linguistics, Santa Fe, N.M., 2018* (Association for Computational Linguistics, 2018), pp. 1650–1661. <https://aclanthology.org/C18-1140>
8. E. I. Bolshakova and N. E. Efremova, "A heuristic strategy for extracting terms from scientific texts," in *Communications in Computer and Information Science*, Ed. by M. Khachay, N. Konstantinova, A. Panchenko, D. Ignatov, and V. Labunets, Communications in Computer and Information Science, Vol. 542 (Springer, Cham, 2015), pp. 297–307. https://doi.org/10.1007/978-3-319-26123-2_29
9. G. Bordea, P. Buitelaar, S. Faralli, and R. Navigli, "SemEval-2015 Task 17: Taxonomy extraction evaluation (TExEval)," in *Proc. 9th Int. Workshop on Semantic Evaluation (SemEval 2015), Denver, Colo., 2016* (Association for Computational Linguistics, 2016), pp. 902–910. <https://doi.org/10.18653/v1/s15-2151>
10. G. Bordea, E. Lefever, and P. Buitelaar, "SemEval-2016 Task 13: Taxonomy extraction evaluation (TExEval-2)," in *Proc. 10th Int. Workshop on Semantic Evaluation (SemEval-2016), San Diego, Calif., 2016* (Association for Computational Linguistics, 2016), pp. 1081–1091. <https://doi.org/10.18653/v1/s16-1168>
11. J. Camacho-Collados, C. Delli Bovi, L. Espinosa Anke, S. Oramas, T. Pasini, E. Santus, V. Shwartz, R. Navigli, and H. Saggion, "SemEval-2018 Task 9: Hypernym discovery," in *Proc. 12th Int. Workshop on Semantic Evaluation, New Orleans, 2018* (Association for Computational Linguistics, 2018), pp. 712–724. <https://doi.org/10.18653/v1/s18-1115>
12. S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. Am. Soc. Inf. Sci.* **41**, 391–407 (1990). [https://doi.org/10.1002/\(sici\)1097-4571\(199009\)41:6<391::aid-asil>3.0.co;2-9](https://doi.org/10.1002/(sici)1097-4571(199009)41:6<391::aid-asil>3.0.co;2-9)
13. J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. 2019 Conf. of the North American Association for Computational Linguistics, Minneapolis, Minn., 2019* (Association for Computational Linguistics, 2019), pp. 4171–4186. <https://doi.org/10.18653/v1/n19-1423>
14. K. Erk, "Vector space models of word meaning and phrase meaning: A survey," *Language Linguist. Compass* **6**, 635–653 (2012). <https://doi.org/10.1002/lnco.362>
15. R. Fu, J. Guo, B. Qin, W. Che, H. Wang, and T. Liu, "Learning semantic hierarchies via word embeddings," in *Proc. 52nd Annu. Meeting of the Assoc. for Computational Linguistics, Baltimore, Md.* (Association for Computational Linguistics, 2014), Vol. 1, pp. 1199–1209. <https://doi.org/10.3115/v1/p14-1113>
16. A. Gómez-Pérez and D. Manzano-Macho, "An overview of methods and tools for ontology learning from texts," *Knowl. Eng. Rev.* **19**, 187–212 (2004). <https://doi.org/10.1017/s0269888905000251>
17. Z. Harris, "Distributional structure," *Word* **10**, 146–162 (1954). <https://doi.org/10.1080/00437956.1954.11659520>
18. M. Hearst, "Automatic acquisition of hyponyms from large text corpora," in *Proc. 14th Conf. on Computational Linguistics, Nantes, 1992* (Association for Computational Linguistics, 1992), Vol. 2, pp. 539–545. <https://doi.org/10.3115/992133.992154>
19. D. Jurgens and M. T. Pilehvar, "SemEval-2016 Task 14: Semantic taxonomy enrichment," in *Proc. 10th Int. Workshop on Semantic Evaluation (SemEval-2016), San Diego, Calif., 2016* (Association for Computational Linguistics, 2016), pp. 1092–1102. <https://doi.org/10.18653/v1/s16-1169>
20. Yu. A. Kiselev, S. V. Porshnev, and M. Yu. Mukhin, "Method of extracting hyponym-hypernym relationships for nouns from definitions of explanatory dictionaries," *Programmnaya Inzh.* **10**, 38–48 (2015).
21. A. Kutuzov and E. Kuzmenko, "WebVectors: A toolkit for building web interfaces for vector semantic models," in *Communications in Computer and Information Science*, Ed. by D. I. Ignatov et al., Communications in Computer and Information Science, Vol. 661 (Springer, Cham, 2017), pp. 155–161. https://doi.org/10.1007/978-3-319-52920-2_15
22. A. Lenci, "Distributional semantics in linguistic and cognitive research," *Italian J. Linguist.* **20** (1), 1–31 (2008).
23. O. Levy, S. Remus, C. Biemann, and I. Dagan, "Do supervised distributional methods really learn lexical inference relations?," in *Proc. 2015 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, Colo., 2015* (Association for Computational Linguistics, 2015), pp. 970–976. <https://doi.org/10.3115/v1/n15-1098>
24. N. V. Loukachevitch, A. A. Gerasimova, B. B. Dobrov, G. Lashevich, and V. V. Ivanov, "Creating Russian WordNet by conversion," in *Computational Linguistics and Intellectual Technologies: Proc. Int. Conf. Dialogue 2016, Moscow, 2016* (2016), pp. 22–30.
25. T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. 26th Int. Conf. on Neural Information Processing Systems, Lake Tahoe, Nevada, 2013*, Ed. by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Curran Associates, Red Hook, N.Y., 2013), pp. 3111–3119.
26. G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM* **38** (11), 39–41 (1995). <https://doi.org/10.1145/219717.219748>
27. N. Nakashole, G. Weikum, and F. Suchanek, "PATTY: A taxonomy of relational patterns with semantic types," in *Proc. 2012 Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, Korea, 2012* (Association for Computational Linguistics, 2012), pp. 1135–1145.

28. I. Nikishina, V. Logacheva, A. Panchenko, and N. Loukachevitch, "RUSSE'2020: Findings of the first taxonomy enrichment task for the Russian language," in *Computational Linguistics and Intellectual Technologies: Proc. Int. Conf. Dialogue 2020* (Russian State University for the Humanities, Moscow, 2020). <https://doi.org/10.28995/2075-7182-2020-19-579-595>
29. I. Nikishina, V. Logacheva, A. Panchenko, and N. Loukachevitch, "Studying taxonomy enrichment on diachronic WordNet versions," in *Proc. 28th Int. Conf. on Computational Linguistics, Barcelona, 2020* (International Committee on Computational Linguistics, 2020), pp. 3095–3106. <https://doi.org/10.18653/v1/2020.coling-main.276>
30. I. Nikishina, M. Tikhomirov, V. Logacheva, Yu. Nazarov, A. Panchenko, and N. Loukachevitch, "Taxonomy enrichment with text and graph vector representations," *Semantic Web* **13**, 441–475 (2022). <https://doi.org/10.3233/sw-212955>
31. S. Roller, K. Erk, and G. Boleda, "Inclusive yet selective: Supervised distributional hypernymy detection," in *Proc. COLING 2014, the 25th Int. Conf. on Computational Linguistics: Technical Papers, Dublin, 2014* (Dublin City Univ. and Association for Computational Linguistics, 2014), pp. 1025–1036.
32. S. Roller, D. Kiela, and M. Nickel, "Hearst patterns revisited: Automatic hypernym detection from large text corpora," in *Proc. 56th Annu. Meeting of the Association for Computational Linguistics, Melbourne, 2018* (Association for Computational Linguistics, 2018), Vol. 2, pp. 358–363. <https://doi.org/10.18653/v1/p18-2057>
33. K. Sabirova and A. Lukanin, "Automatic extraction of hypernyms and hyponyms from Russian texts," *CEUR Workshop Proc.* **1197**, 35–40 (2014). <https://ceur-ws.org/Vol-1197/paper6.pdf>
34. T. Schick and H. Schütze, "Exploiting cloze-questions for few-shot text classification and natural language inference," in *Proc. 16th Conf. of the European Chapter of the Association for Computational Linguistics: Main Volume* (Association for Computational Linguistics, 2021), pp. 255–269. <https://doi.org/10.18653/v1/2021.eacl-main.20>
35. V. Schwartz, Yo. Goldberg, and I. Dagan, "Improving Hypernymy Detection with an Integrated Path-based and Distributional Method," in *Proc. 54th Annu. Meeting of the Association for Computational Linguistics, Berlin, 2016* (Association for Computational Linguistics, 2016), Vol. 1, pp. 2389–2398. <https://doi.org/10.18653/v1/p16-1226>
36. V. Schwartz, E. Santus, and D. Schlechtweg, "Hypernyms under siege: Linguistically-motivated artillery for hypernymy detection," in *Proc. 15th Conf. of the European Chapter of the Association for Computational Linguistics, Valencia, 2017* (Association for Computational Linguistics, 2017), Vol. 1, pp. 65–75. <https://doi.org/10.18653/v1/e17-1007>
37. J. Seitner, C. Bizer, K. Eckert, S. Faralli, R. Meusel, H. Paulheim, and S. Ponzetto, "A large database of hypernymy relations extracted from the web," in *Proc. Tenth Int. Conf. on Language Resources and Evaluation (LREC'16), Portorož, Slovenia, 2016* (European Language Resources Association, 2016), pp. 360–367. <https://aclanthology.org/L16-1056>
38. J. Shen, Z. Shen, C. Xiong, C. Wang, K. Wang, and J. Han, "TaxoExpan: Self-supervised taxonomy expansion with position-enhanced graph neural network," in *Proceedings of The Web Conference 2020, Taipei, 2020*, Ed. by Ye. Huang, I. King, T.-Ya. Liu, and M. van Steen (Association for Computing Machinery, New York, 2020), pp. 486–497. <https://doi.org/10.1145/3366423.3380132>
39. R. Snow, D. Jurafsky, and A. Y. Ng, "Learning syntactic patterns for automatic hypernym discovery," in *Proc. 17th Int. Conf. on Neural Information Processing, Vancouver, 2004*, Ed. by L. K. Saul, Y. Weiss, and L. Bottou (MIT Press, Cambridge, Mass., 2004), pp. 1297–1304. <https://doi.org/10.7551/mitpress/7503.003.0111>
40. R. Snow, D. Jurafsky, and A. Y. Ng, "Semantic taxonomy induction from heterogenous evidence," in *Proc. 21st Int. Conf. on Computational Linguistics and the 44th Annu. Meeting of the ACL, Sydney, 2006* (Association for Computational Linguistics, Stroudsburg, Pa., 2006), pp. 801–808. <https://doi.org/10.3115/1220175.1220276>
41. F. Tian, C. Yuan, and F. Ren, "Hyponym extraction from the web by bootstrapping," *IEEE Trans. Electr. Electron. Eng.* **7** (1), 62–68 (2012). <https://doi.org/10.1002/tee.21696>
42. M. Tikhomirov and N. Loukachevitch, "Meta-embeddings in taxonomy enrichment task," in *Computational Linguistics and Intellectual Technologies: Papers from the Annual Conf. Dialogue-2021* (Russian State University for the Humanities, Moscow, 2021), pp. 681–691. <https://doi.org/10.28995/2075-7182-2021-20-681-691>
43. P. D. Turney and P. Pantel, "From frequency to meaning: Vector space models of semantics," *J. Artif. Intell. Res.* **37**, 141–188 (2010). <https://doi.org/10.1613/jair.2934>
44. D. Ustalov, N. Arefyev, C. Biemann, and A. Panchenko, "Negative sampling improves hypernymy extraction based on projection learning," in *Proc. 15th Conf. of the Eur. Chapter of the Assoc. for Computational Linguistics, Valencia, 2017* (Association for Computational Linguistics, 2017), Vol. 2, p. 543. <https://doi.org/10.18653/v1/e17-2087>
45. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Int. Conf. on Neural Information Processing, Long Beach, Calif., 2017*, Ed. by U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, and R. Fergus (Curran Associates, Red Hook, N.Y., 2017), pp. 6000–6010.
46. H. Xu, Yu. Chen, Z. Liu, Ya. Wen, and X. Yuan, "TaxoPrompt: A prompt-based generation method with taxonomic context for self-supervised taxonomy expansion," in *Proc. Thirty-First Int. Joint Conf. on Artificial Intelligence*, Ed. by L. De Raedt (International Joint Conferences on Artificial Intelligence Organization, 2022), pp. 4432–4438. <https://doi.org/10.24963/ijcai.2022/615>

47. J. Yamane, T. Takatani, H. Yamada, M. Miwa, and Yu. Sasaki, "Distributional hypernym generation by jointly learning clusters and projections," in *Proc. COLING 2016, the 26th Int. Conf. on Computational Linguistics: Technical Papers, Osaka, Japan, 2016* (The COLING 2016 Organizing Committee, 2016), pp. 1871–1879. <https://aclanthology.org/C16-1176>.
48. H. Yang and J. Callan, "A metric-based framework for automatic taxonomy induction," in *Proc. Joint Conf. of the 47th Annu. Meeting of the ACL and the 4th Int. Joint Conf. on Natural Language Processing of the AFNLP, Suntec, Singapore, 2009* (Association for Computational Linguistics, Stroudsburg, Pa., 2009), Vol. 1, pp. 271–279. <https://doi.org/10.3115/1687878.1687918>

Translated by L.A. Solovyova



Natalya V. Loukachevitch. Born in 1964. Graduated from the Faculty of Computational Mathematics and Cybernetics of Moscow State University; defended her thesis for the degree of Doctor of Technical Sciences in 2016. Currently, she is a leading researcher at the Research Computing Center of the Moscow State University. She lectures on automatic text processing and information retrieval at the Faculty of Computational Mathematics and Cybernetics and the Faculty of Philology of Moscow State University, as well as at Bauman Moscow State Technical University. She has more than 300 publications in the field of automatic word processing, information retrieval, and knowledge representation. Research interests: automatic text processing, information retrieval, ontologies.