Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2017"

Moscow, May 31—June 3, 2017

EXPANDING HIERARCHICAL CONTEXTS FOR CONSTRUCTING A SEMANTIC WORD NETWORK

Ustalov D. A. (dau@imm.uran.ru)

Krasovskii Institute of Mathematics and Mechanics;
Ural Federal University, Yekaterinburg, Russia

A semantic word network is a network that represents the semantic relations between individual words or their lexical senses. This paper proposes Watlink, an unsupervised method for inducing a semantic word network (SWN) by constructing and expanding the hierarchical contexts using both the available dictionary resources and distributional semantics' methods for *is-a* relations. It has three steps: context construction, context expansion, and context disambiguation. The proposed method has been evaluated on two different datasets for the Russian language. The former is a well-known lexical ontology built by the group of expert lexicographers. The latter, LRWC ("Lexical Relations from the Wisdom of the Crowd"), is a new resource created using crowdsourcing that contains both positive and negative human judgements for subsumptions. The proposed method outperformed the other relation extraction methods on both datasets according to recall and F_1-score. Both the implementation of the Watlink method and the LRWC dataset are publicly available under libré licenses.

Keywords: lexical semantics, hyponym, hypernym, subsumption, semantic network, crowdsourcing, Russian

ПОСТРОЕНИЕ СЕМАНТИЧЕСКОЙ СЕТИ СЛОВ ПУТЁМ РАСШИРЕНИЯ ИЕРАРХИЧЕСКИХ КОНТЕКСТОВ

Усталов Д. А. (dau@imm.uran.ru)

Институт математики и механики им. Н.Н.Красовского Уральского отделения РАН; Уральский федеральный университет, Екатеринбург, Россия

Семантическая сеть слов — это сеть, представляющая семантические отношения между отдельными словами или значениями слов. В данной работе представлен метод Watlink для построения семантической

сети слов на основе обучения без учителя. Метод включает три этапа: формирование иерархических контекстов, расширение иерархических контекстов, связывание полученных контекстов. Произведена оценка представленного метода на двух разных наборах данных для русского языка: по материалам тезауруса RuWordNet и по материалам нового набора данных LRWC, содержащего суждения людей о родовидовых связях русских слов, полученных при помощи краудсорсинга. Предложенный метод продемонстрировал более высокие значения полноты и F_1-меры на обоих наборах данных по сравнению с другими методами извлечения отношений. Реализация метода Watlink и набор данных LRWC доступны на условиях открытой лицензии.

Ключевые слова: лексическая семантика, гипоним, гипероним, родовидовое отношение, семантическая сеть, краудсорсинг, русский язык

1. Introduction

A semantic network is a network that represents semantic relations between concepts [27]. Such semantic networks as WordNet [8] and BabelNet [21] are successfully applied in addressing different problems requiring common sense reasoning. Construction of such a network 'by hand' is a long and expensive process that involves very large amount of efforts of expert lexicographers. For instance, the Russian language is still considered as an under-resourced natural language [12], which makes it highly topical to develop new methods for discovering and refining the available dictionaries and other lexical semantic resources in an unsupervised way.

This paper is focused on a special kind of semantic networks—semantic word networks—that represent the relations between individual words or their lexical senses rather than the entire concepts [15]. Semantic word networks (SWNs) found their application in marketing campaign optimization [25], search query expansion [11], etc. Particularly, this paper is devoted to the *hyponymy/hypernymy* relation, also known as the *is-a* or the *subsumption* relation. Thus, an SWN is a directed graph connecting the distinct lexical senses through the hypernymy relations.

The contribution of the present paper is two-fold. Firstly, Watlink, an unsupervised method for constructing an SWN that uses both distributional and dictionary resources has been proposed. Secondly, a crowdsourced dataset LRWC ("Lexical Relations from the Wisdom of the Crowd") representing both positive and negative human judgements for hyponymy and hypernymy relations for the Russian language has been disseminated. The proposed method is inspired by the one by Faralli et al. [7]. The difference is that the present method disambiguates synsets and their hierarchical contexts instead of distributional sense representations. Also, it provides an optional context expansion step to increase the lexical coverage of the resulting dataset.

The rest of the paper is organized as follows. Section 2 reviews the related work focused on the construction of *is-a* relations. Section 3 describes Watlink, an unsupervised method for semantic word network construction. Section 4 shows the evaluation results of this method on a well-known gold standard dataset for Russian. Section 5 presents the LRWC dataset and demonstrates the evaluation results on this new dataset. Section 6 concludes with the final remarks.

2. Related Work

Currently, the most widely used method for detecting hypernyms and hyponyms is the Hearst patterns [10]. These lexical-syntactic patterns, e.g., "Y such as X_1 and X_2", have successfully found a substantial number of applications including ontology learning. There is a couple of variations of such patterns like PatternSim [22] and sense definition parsing [13], but the core principle remains the same and these patterns suffer from the sparsity problem.

Various forms of crowdsourcing are used for constructing or expanding lexical resources. Wiktionary, a wiki-based dictionary, is a popular source of semantic information [31]. Also, there are other initiatives like the BabelNet Annotation Group (BANG) [21] and Yet Another RussNet (YARN) [4], which differ in the goals and deliverables.

Distributed word representations, also known as word embeddings [20], are a trending topic nowadays. Fu et al. [9] proposed the projection learning approach for constructing semantic hierarchies for the Chinese language. This approach assumes learning a linear transformation matrix such that multiplying on which a hyponym vector produces a hypernym vector. Also, the k-means clustering algorithm has been used to split the embeddings space into several subspaces to provide more flexibility to the model. Recently, this approach has been improved by negative sampling, which yielded a significant quality boost [28].

Shwartz et al. developed HypeNET, an integrated method that combines the syntactic parsing features with word embeddings based on a long short-term memory network [26]. However, HypeNET requires high-quality dependency pairs, which complicates its application for under-resourced languages.

3. Constructing a Semantic Word Network with Watlink

Let $\mathcal{S}$ be the input set of synsets, and a synset $S \in \mathcal{S}$ is a set of semantically equivalent word senses[1], e.g., $\{auto^2, car^1, automobile^1, \ldots\}$. Let $\mathcal{R}$ be the input set of *is-a* relations provided in the form of tuples $(w, h) \in \mathcal{R}$, where both the hyponym w and the hypernym h have no sense labels attached, e.g., (*bank*, *building*). The goal is to assign the corresponding sense labels to these relations as well as to provide the words with missing hypernyms with those, if possible.

For that, the Watlink method, shown in Fig. 1, is proposed. Firstly, a hierarchical context representing a bag of hypernym words is constructed for each synset. Secondly, each hierarchical context is expanded using the nearest neighbor retrieval combined with projection learning. Finally, the sense labels for the hypernyms are obtained using the context disambiguation. As the result, Watlink constructs a directed graph SWN $= (V, E)$, where $V = \bigcup_{S \in \mathcal{S}} S$ is the set of all the possible word senses appearing in all the synsets, and $E \subseteq V \times V$ is the set of disambiguated *is-a* relations between these senses.

[1] Following the notation used in BabelNet [21], $word^i$ denotes the i-th lexical sense of the given *word*.

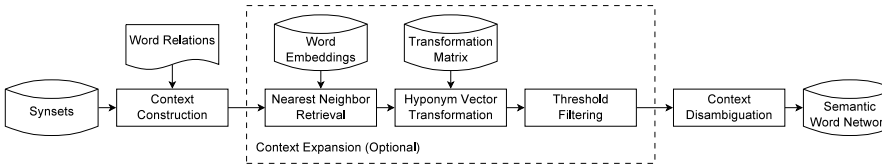

Fig. 1. An outline of the proposed method, WATLINK

3.1. Context Construction

A hierarchical context connects the lexical entries of the corresponding synset to their typical hypernyms in the *is-a* dictionary. For a synset $S \in \mathcal{S}$, the hierarchical context $\mathrm{hctx}(S)$ is a bag of words composed of all the hypernyms in $\mathcal{R}$ matching the words in S as hyponyms:

$$\mathrm{hctx}(S) = \{h : w \in \mathrm{words}(S), (w, h) \in \mathcal{R}\},$$

where $\mathrm{words}(S)$ is the set of lemmas corresponding to the word senses in S, (w, h) is a pair of hyponym w and hypernym h present in the dictionary $\mathcal{R}$. As the result, hypernyms are propagated to the words in the synset for which no hypernyms were available. The variable importance of words in hierarchical contexts is modeled using tf, idf, and tf-idf [19]:

$$\text{tf-idf}(h, \mathrm{hctx}(S), \mathcal{S}) = \underbrace{\frac{|h' \in \mathrm{hctx}(S) : h = h'|}{|\mathrm{hctx}(S)|}}_{\mathrm{tf}(h,\, \mathrm{hctx}(S))} \times \underbrace{\log\left(\frac{|\mathcal{S}|}{|S' \in \mathcal{S} : h \in \mathrm{hctx}(S')|}\right)}_{\mathrm{idf}(h, \mathcal{S})}$$

For example, a hierarchical context for the synset mentioned in the beginning of Section 3, can be like {*vehicle, transport, motor vehicle*} for $\mathrm{words}(S)$ = {*auto*, *car*, *automobile*}.

3.2. Context Expansion

Given the fact that the amount of available resources representing *is-a* relations is limited [12], a projection learning approach [9] has been used to expand the set of the already available subsumption pairs $(w, h) \in \mathcal{R}$. This optional context expansion step is based on searching the most similar words to h using the semantic similarity computed on word embeddings [20] and filtering out the irrelevant candidates (Fig. 2).

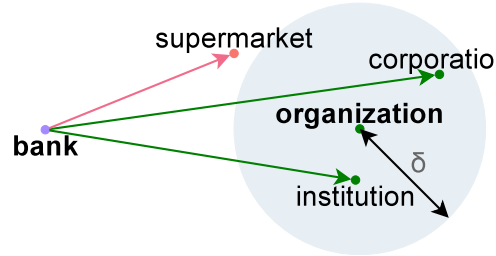

Fig. 2. Hypernyms of "bank" that are similar to a known hypernym "organization": the candidate words "corporation" and "institution" within a radius of δ are correct, while the candidate word "supermarket" is not

Firstly, for each input *is-a* pair $(w, h) \in \mathcal{R}$, a set of n nearest neighbors $\mathrm{NN}_n(\vec{h})$ of the hypernym embedding $\vec{h}$ is retrieved. Secondly, for each hypernym candidate embedding $\vec{h'} \in \mathrm{NN}_n(\vec{h})$, a transformation matrix Φ^*, corresponding to the subspace of the vector offset $\vec{h'} - \vec{w}$, is obtained and multiplied on the hyponym embedding $\vec{w}$ [9], resulting in the predicted hypernym vector $\Phi^*\vec{w}$. Finally, the Euclidean distance between $\Phi^*\vec{w}$ and the hypernym embedding $\vec{h}$ is computed. Those predicted vectors located within a radius of δ from the latter vector are said to be relevant: $\left\|\Phi^*\vec{w} - \vec{h}\right\| < \delta$. As the result, a hierarchical context $\mathrm{hctx}(S)$ can be transformed into the expanded hierarchical context $\mathrm{hctx}'(S)$:

$$\mathrm{hctx}'(S) = \bigcup_{\substack{w \in \mathrm{words}(S),\\ (w,h) \in \mathcal{R}}} \{w\} \times \mathrm{NN}_n^*(\vec{h}) \cup \mathrm{hctx}(S)$$

where $\vec{w}$ and $\vec{h}$ are embedding vectors for the words w and h, correspondingly, $\mathrm{NN}_n^*(\vec{h})$ is the set of relevant candidates of n nearest neighbors of the vector $\vec{h}$.

3.3. Context Disambiguation

For each synset $S \in \mathcal{S}$ and its hierarchical context $\mathrm{hctx}(S)$, a sense label is estimated for each hypernym $h \in \mathrm{hctx}(S)$. This is achieved by selecting a synset $S' \in \mathcal{S}: h \in \mathrm{words}(S')$ that maximizes the cosine similarity [7] between $\mathrm{hctx}(S)$ and S to choose the optimal word sense $\hat{h}$:

$$\hat{h} = \underset{\substack{S' \in \mathcal{S}, S \neq S', h' \in S',\\ \mathrm{words}(\{h'\}) = h}}{\arg\max} \cos(\mathrm{hctx}(S),\ S')$$

For instance, consider the hierarchical context $\{material, data\}$ and two synsets: $\{material^1, textile^1\}$ and $\{material^2, information^1, data^1\}$. Using this procedure, the second sense of the word "material" will be chosen because the latter synset is more similar to the given hierarchical context. The resulting disambiguated hierarchical context contains the sense labels attached to the words composing the initial hierarchical context, i.e., $\widehat{hctx}(S) = \{\hat{h} : h \in \mathrm{hctx}(S)\}$. It is now possible to construct an SWN with $\bigcup_{S \in \mathcal{S}} S$ as the set of nodes and $\bigcup_{S \in \mathcal{S}} S \times \widehat{hctx}(S)$ as the set of edges that are labeled *is-a* relations. An example of an SWN is presented in Fig. 3. Note that the ambiguous word "ticket" is represented twice: $ticket^1$ is a document and $ticket^2$ is a sign.

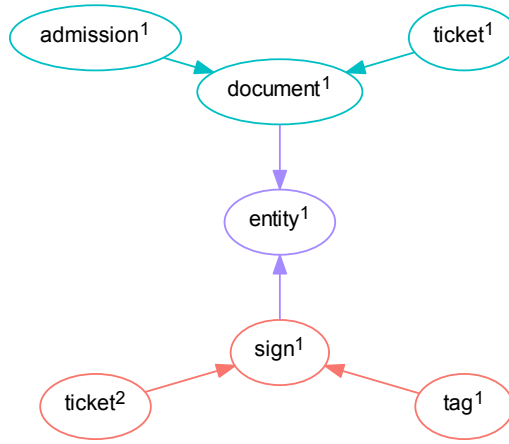

Fig. 3. An example of a semantic word network

4. Gold Standard Evaluation on RuWordNet

Since Watlink implies no strict limitations on the structure of the input synsets and *is-a* relations, it can be applied for linking senses in virtually any synset dataset, e.g., YARN [4] or UNLDC [5], with the relations from practically any subsumption dataset, e.g., Hearst patterns [10], Wiktionary [31], etc. During the gold standard evaluation, the performance of the Watlink method is studied on RuWordNet [17], a WordNet-like version of the RuThes thesaurus for Russian [16].

4.1. Experimental Setup

Given the fact that Watlink is an unsupervised method, except the optional expansion step, the synset dataset is also chosen to be obtained in an unsupervised way. A state-of-the-art method for unsupervised synset induction, Watset [29], has been used to yield the synsets from the synonymy graph composed of three synonymy dictionaries for Russian: the Russian Wiktionary [31], UNLDC [5] and the Abramov's Dictionary [1]. This resulted in 55,369 synsets uniting 83,092 lexical entries; Watset was configured to use the Chinese Whispers algorithm for word sense induction [2] and the Markov cluster algorithm for global graph clustering [30].

The following *is-a* datasets have been used in the experiments:

- *Patterns*, the dataset extracted from the lib.rus.ec electronic library using the PatternSim approach [22]; the *Limit* option specifies that only these relations appeared at least $f = 30$ times have remained.
- *Wiktionary*, the dataset extracted from the Russian Wiktionary using the JWKTL tool [31].
- *SAD* the dataset extracted from the sense definitions in the Small Academic Dictionary [6] [13].
- *Joint*, the dataset uniting *Patterns* + *Limit*, *Wiktionary* and *SAD*.

Each dataset has been expanded using the context expansion method described in Section 3.2 with $n = 10$, which is denoted as *Exp*. The transformation matrices have been estimated using projection learning with asymmetric regularization [28] on the state-of-the-art 500-dimensional skip-gram word embeddings for Russian [23]. The threshold $\delta = 0.6$ has been tuned separately on a development dataset.

4.2. Evaluation Metric

The performance is reported according to the pairwise information retrieval quality metrics: precision, recall and F_1-score [19]. For that. an *is-a* pair (*hypo*, *hyper*) is considered as predicted correctly if and only if there is a path from some sense of *hypo* to some sense of *hyper* exists in the gold standard dataset. Only the words appearing both in the gold standard and the comparable datasets are considered. The rest words are excluded from the evaluation.

4.3. Results

In most cases, the tf-idf weighing approach yielded slightly better results according to F_1-score (Fig. 4). Table 1 summarize the evaluation results obtained on the RuWordNet dataset using tf-idf weights. The highlighted results are statistically significant according to the Wilcoxon signed-rank test with the significance level of 0.01 performed similarly to [24]. It clearly seems that the expansion increases recall with a slight, yet notable, drop of precision. However, SWN outperformed the others in terms of recall and F_1-score.

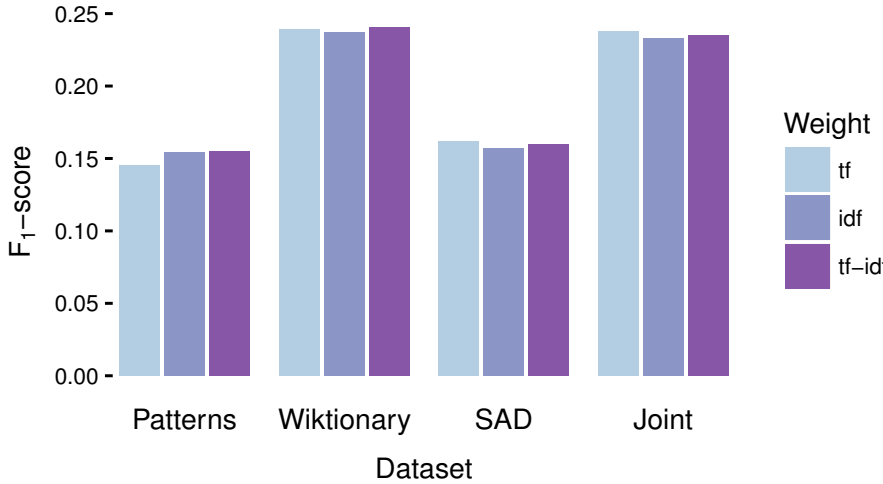

Fig. 4. Influence of the weighing approach according to the best result on each dataset

Table 1. Evaluation results on the RuWordNet dataset, only the best configurations shown, top three results are highlighted

Method	# of pairs	Precision	Recall	F_1-score
Patterns	1,597,651	0.1611	0.3255	0.2155
Patterns + SWN	236,922	0.1126	0.2451	0.1543
Patterns + Limit	10,458	**0.3773**	0.0157	0.0302
Patterns + Limit + Exp	10,715	0.3760	0.0160	0.0307
Patterns + Limit + SWN	46,758	0.1140	0.0717	0.0880
Patterns + Limit + Exp + SWN	47,387	0.1129	0.0722	0.0881
Wiktionary	108,985	**0.3877**	0.0898	0.1458
Wiktionary + Exp	110,329	**0.3874**	0.0907	0.1469
Wiktionary + SWN	177,787	0.1836	0.3460	**0.2399**
Wiktionary + Exp + SWN	179,623	0.1844	**0.3464**	**0.2407**
SAD	36,800	0.1823	0.1502	0.1647
SAD + Exp	37,702	0.1825	0.1515	0.1655
SAD + SWN	98,085	0.1383	0.1879	0.1593
SAD + Exp + SWN	99,678	0.1385	0.1883	0.1596
Joint	149,195	0.1719	0.2590	0.2067
Joint + Exp	151,150	0.1720	0.2594	0.2069
Joint + SWN	216,285	0.1685	**0.3865**	0.2347
Joint + Exp + SWN	218,290	0.1687	**0.3867**	**0.2350**

5. Lexical Relations from the Wisdom of the Crowd

To study the performance of the proposed method more thoroughly, the best models in Section 4.3 have been chosen as the input subsumption pairs for collecting human judgements.

A set of 300 most frequent nouns have been extracted from the Russian National Corpus [18]. Then, each method or resource in Table 1, produced at most five hypernyms for each of these 300 nouns, if possible. In case it is not possible, missing answers treated as false negative answers. Two additional datasets participated in the evaluation: RuThes [16] and a noun-only version of RuWordNet [17]. The order of the extracted hypernyms is the same as in they are presented in the resource. This resulted in 10,600 unique non-empty subsumption pairs that have been passed for crowdsourcing annotation on the Yandex.Toloka[2] platform. Each pair has been annotated by seven different annotators whose mother tongue is Russian and the age is at least 20 by February 1, 2017.

Prior to this annotation, a manually composed training set of 48 tasks for less frequent nouns has been ran. Only those who answered correctly for at least 80 % of the training tasks have been permitted to complete non-training tasks for paid. Also, the workers have been provided with a detailed instruction containing recommendations among the examples of correct positive and negative answers.

[2] https://toloka.yandex.ru/

5.1. Human Intelligence Task

The layout of the human intelligence task (HIT) design, depicted in Fig. 5, assumes the direct answer to a simple question: does the given pair of words represent a meaningful *is-a* relation? Since the crowd workers are not expert lexicographers and this question might be difficult for them, it has been rephrased as "Is it correct that a kitten is a kind of mammal?" (in Russian).

Правда ли, что **банк** — это разновидность **организации**?
◯ Да
◯ Нет

Fig. 5. Layout of the HIT: "Is it correct that a *bank* is a kind of *organization*?" (Yes / No), both the hyponym and hypernym words link to the Yandex search page; note that the hypernym is represented in the genitive case: «организации» instead of «организация»

In case of English, it will be sufficient to provide just the lemmas for both the hyponym and the hypernym. In Russian, this will make the question sentence uncoordinated because the hypernym word should be present in the genitive case. Also, such words as «молоко» (milk) and «дом» (house) are written identically both in nominative and accusative cases, which causes inflection problems. This limitation has been dealt with the pymorphy2 morphological analyzer and generator [14] by estimating the most suitable word form that needs to be inflected into the genitive case according to the heuristic

$$\text{score}(w|t,c) = \text{p}(t|w) + 1(t = \text{noun}) \times 10 + 1(c = \text{nominative}) \times 2$$

where $1(\cdot)$ is the indicator function, $\text{p}(t|w)$ is the probability of the tag t assigned to the word w estimated on OpenCorpora [3], and c is the grammatical case. This heuristic prefers the nouns in the nominative case because the input words in the present study are in fact noun lemmas.

5.2. Dataset

The answers have been aggregated using the Yandex.Toloka proprietary answer aggregation mechanism. As the result, 4,576 out of 10,600 pairs have been annotated as positive while the rest 6,024 have been annotated as negative. Interestingly, in average, the workers were more confident in negative answers rather than in the positive ones according to the two-tailed t-test with the significance level of 0.01. These negative answers are extremely useful for both training and testing different relation extraction methods [28]. To the best of our knowledge, this is the first dataset of this kind made for the Russian language using microtask-based crowdsourcing.

5.3. Experimental Setup

Since for each of the top 300 nouns each method should provide no less and no more than five hypernyms, including the missing ones, the performance of the

methods is quantitated using the precision, recall, and F_1-score. A hypernym is considered as predicted correctly if and only if it is not empty and is annotated as positive by the crowd workers.

5.4. Results

The evaluation results on LRWC showed that this resource does correlate with RuThes in terms of correctness. However, it is not necessarily a reliable source of subsumptions given the fact that it represents the human judgements, not the expert knowledge.

Table 2. Evaluation results on the LRWC dataset, top three results are highlighted

Method	Precision	Recall	F_1-score
RuThes	**0.7035**	**0.9168**	**0.7961**
Joint + Exp	0.6719	0.9002	**0.7695**
Joint	0.6726	0.8975	**0.7690**
Wiktionary + SWN	0.6287	0.8775	0.7326
Wiktionary + Exp + SWN	0.6254	0.8779	0.7304
Joint + SWN	0.5590	**0.9306**	0.6985
Joint + Exp + SWN	0.5569	**0.9304**	0.6968
RWN (Nouns)	0.5878	0.8400	0.6917
SAD + Exp	0.6313	0.6141	0.6226
SAD	0.6321	0.6121	0.6220
Patterns	0.4821	0.8710	0.6207
Wiktionary + Exp	**0.7488**	0.3485	0.4756
Wiktionary	**0.7492**	0.3467	0.4741
Patterns + Limit	0.6711	0.3103	0.4244
Patterns + Limit + Exp	0.6700	0.3105	0.4244

RuThes showed the best results on the LRWC dataset according to F1-score and the third best result according to precision and recall. Surprisingly, the highest value of precision is yielded by the Wiktionary dataset that has been created using crowdsourcing by a group of volunteers. Although it shows relatively small recall, this observation indicates a tremendous potential of collaborative lexicography.

With expansion or without it, SWN showed the designed trade-off between precision and recall in the favor of recall, which agrees with the previous experiment (Table 1). It leaves one with a choice: it is possible to either achieve the highest recall by ignoring the information encoded by the synonyms and their common hypernyms, or to exploit this information by slightly reducing the recall while maintaining the third best value of F_1-score. Notably, on LRWC, the Joint dataset yielded better results without SWN. This is caused by the hypernymy propagation property of the method mentioned in Section 3.1, i.e., when the most frequent hypernyms overweight the others in a hierarchical context.

6. Conclusion

In this paper, Watlink, a robust method for constructing a semantic word network has been proposed. The method showed good results by outperforming competing methods on recall and F_1-score on two different datasets: an expert-built thesaurus, RuWordNet, and a new dataset representing the human judgments for Russian subsumptions, LRWC.

The implementation of the present method is available on GitHub under the MIT license: https://github.com/dustalov/watlink. The LRWC dataset is available on Zenodo in the tab-separated values format under a Creative Commons Attribution-ShareAlike license: https://doi.org/10.5281/zenodo.546302.

Acknowledgements. The reported study is funded by RFBR according to grant no. 16-37-00354 мол_а. The author is grateful to Natalia Loukachevitch who provided machine-readable versions of the RuThes and RuWordNet datasets. The author would also like to thank Alexander Panchenko, Nikolay Arefyev, Andrey Sozykin, and three anonymous reviewers for useful comments on the present study.

References

1. *Abramov N.* (1999), The dictionary of Russian synonyms and semantically related expressions [Slovar' russkikh sinonimov i skhodnykh po smyslu vyrazhenii], Russkie Slovari, Moscow, Russia.
2. *Biemann C.* (2006), Chinese Whispers: An Efficient Graph Clustering Algorithm and Its Application to Natural Language Processing Problems, Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing, New York, NY, USA, pp. 73–80.
3. *Bocharov V. V., Alexeeva S. V., Granovsky D. V., Protopopova E. V., Stepanova M. E., Surikov A. V.* (2013), Crowdsourcing morphological annotation, Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", Bekasovo, Russia, pp. 109–124.
4. *Braslavski P., Ustalov D., Mukhin M., Kiselev Y.* (2016), YARN: Spinning-in-Progress, Proceedings of the 8th Global WordNet Conference, Bucharest, Romania, pp. 56–65.
5. *Dikonov V. G.* (2013), Development of lexical basis for the Universal Dictionary of UNL Concepts, Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", Bekasovo, Russia, pp. 212–221.
6. *Evgen'eva A.P.* (1999), Small Academic Dictionary [Malyi akademicheskii slovar'], Rus. yaz; Poligrafresursy, Moscow, Russia.
7. *Faralli S., Panchenko A., Biemann C., Ponzetto S. P.* (2016), Linked Disambiguated Distributional Semantic Networks, The Semantic Web – ISWC 2016: 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part II, pp. 56–64.
8. *Fellbaum C.* (1998), WordNet: An Electronic Lexical Database, MIT Press, Cambridge, MA, USA.

9. *Fu R., Guo J., Qin B., Che W., Wang H., Liu T.* (2014), Learning Semantic Hierarchies via Word Embeddings, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MA, USA, pp. 1199–1209.
10. *Hearst M. A.* (1992), Automatic Acquisition of Hyponyms from Large Text Corpora, Proceedings of the 14th Conference on Computational Linguistics — Volume 2, Nantes, France, pp. 539–545.
11. *Kan K.-L. and Hsueh H.-Y.* (2013), Conceptual Information Retrieval System Based on Automatically Constructed Semantic Word Network, Intelligent Technologies and Engineering Systems, Proceedings of the 2nd International Conference on Intelligent Technologies and Engineering Systems (ICITES2013), Kaohsiung, Taiwan, pp. 277–283.
12. *Kiselev Y., Porshnev S. V., Mukhin M.* (2015), Current Status of Russian Electronic Thesauri: Quality, Completeness and Availability [Sovremennoe sostoyanie elektronnykh tezaurusov russkogo yazyka: kachestvo, polnota i dostupnost'], Software Engineering [Programmnaya inzheneriya], Vol. 6, pp. 34–40.
13. *Kiselev Y., Porshnev S. V., Mukhin M.* (2015), Method of Extracting Hyponym-Hypernym Relationships for Nouns from Definitions of Explanatory Dictionaries [Metod izvlecheniya rodovidovykh otnoshenii mezhdu sushchestvitel'nymi iz opredelenii tolkovykh slovarei], Software Engineering [Programmnaya inzheneriya], Vol. 10, pp. 38–48.
14. *Korobov M.* (2015), Morphological Analyzer and Generator for Russian and Ukrainian Languages, Analysis of Images, Social Networks and Texts: 4th International Conference, AIST 2015, Revised Selected Papers, Yekaterinburg, Russia, pp. 320–332.
15. *Lee S., Lee M., Kim P., Jung H., Sung W.-K.* (2010), OntoFrame S3: Semantic Web-Based Academic Research Information Portal Service Empowered by STARWIN, The Semantic Web: Research and Applications: 7th Extended Semantic Web Conference, ESWC 2010, May 30—June 3, 2010, Proceedings, Part II, Heraklion, Crete, Greece, pp. 401–405.
16. *Loukachevitch N. V.* (2011), Thesauri in Information Retrieval Tasks [Tezaurusy v zadachakh informatsionnogo poiska], Idz-vo MGU, Moscow, Russia.
17. *Loukachevitch N. V., Lashevich G., Gerasimova A. A., Ivanov V. V., Dobrov B. V.* (2016), Creating Russian WordNet by Conversion, Computational Linguistics and Intellectual Technologies: papers from the Annual conference "Dialogue", Moscow, Russia, pp. 405–415.
18. *Lyashevskaya O., Sharoff S.* (2009), Frequency dictionary of modern Russian based on the Russian National Corpus [Chastotnyi slovar' sovremennogo russkogo yazyka (na materialakh Natsional'nogo korpusa russkogo yazyka)], Azbukovnik, Moscow, Russia.
19. *Manning C. D., Raghavan P., Schütze P.* (2008), Introduction to Information Retrieval, Cambridge University Press, Cambridge, UK.
20. *Mikolov T., Sutskever I., Chen K., Corrado G. S., Dean J.* (2013), Distributed Representations of Words and Phrases and their Compositionality, Advances in Neural Information Processing Systems 26, Harrahs and Harveys, NV, USA, pp. 3111–3119.

21. *Navigli R., Ponzetto S. P.* (2012), BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network, Artificial Intelligence, Vol. 193, pp. 217–250.
22. *Panchenko A., Morozova O., Naets, H.* (2012), A Semantic Similarity Measure Based on Lexico-Syntactic Patterns, Proceedings of KONVENS 2012, Vienna, Austria, pp. 174–178.
23. *Panchenko A., Ustalov D., Arefyev N., Paperno D., Konstantinova N., Loukachevitch N., Biemann C.* (2017), Human and Machine Judgements for Russian Semantic Relatedness, Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016, Revised Selected Papers, Yekaterinburg, Russia, pp. 303–317.
24. *Riedl M., Biemann C.* (2016), Unsupervised Compound Splitting With Distributional Semantics Rivals Supervised Methods, Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, pp. 617–622.
25. *Sanchez-Monzon J., Putzke J., Fischbach K.* (2011), Automatic Generation of Product Association Networks Using Latent Dirichlet Allocation, Procedia — Social and Behavioral Sciences, Vol. 26, pp. 63–75.
26. *Shwartz V., Goldberg Y., Dagan, I.* (2016), Improving Hypernymy Detection with an Integrated Path-based and Distributional Method, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, pp. 2389–2398.
27. *Sowa J. F.* (1987), Semantic Networks, available at: http://www.jfsowa.com/pubs/semnet.htm
28. *Ustalov D., Arefyev N., Biemann C., Panchenko A.* (2017), Negative Sampling Improves Hypernymy Extraction Based on Projection Learning, Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, Valencia, Spain, pp. 543–550.
29. *Ustalov D.* (2017), Concept Discovery from Synonymy Graphs [Obnaruzhenie ponyatii v grafe sinonimov], Computational Technologies [Vychislitel'nye tekhnologii], Vol. 22, Special Issue 1, pp. 99–112.
30. *van Dongen S.* (2000), Graph Clustering by Flow Simulation, Ph.D. thesis, University of Utrecht, Utrecht, The Netherlands.
31. *Zesch T., Müller C., Gurevych I.* (2008), Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary, Proceedings of the 6th International Conference on Language Resources and Evaluation, Marrakech, Morocco, pp. 1646–1652.