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

Moscow, May 31—June 3, 2017

A SYNTAX-BASED DISTRIBUTIONAL MODEL FOR DISCRIMINATING BETWEEN SEMANTIC SIMILARITY AND ASSOCIATION

Trofimov I. V. (itrofimov@gmail.com),
Suleymanova E. A. (yes2helen@gmail.com)

Program Systems Institute of RAS, Pereslavl-Zalessky, Russia

In recent years, distributional semantics has shown a trend towards a deeper understanding of what semantic relatedness is and what it is composed of. This is attested, in particular, by the emergence of new gold standards like SimLex999, WS-Sim and WS-Rel. Evidence from cognitive psychology suggests that humans distinguish between two basic types of semantic relations: category-based similarity and thematic association. The paper presents a distributional model capable of differentiating between these relations, and a dataset consisting of 500 similar and 500 associated pairs of nouns that can be used for evaluation of such models.

Keywords: semantic relatedness, semantic similarity, association, RUSSE, RuSim1000 dataset, syntactic-based distributive semantic model, RuWac, MaltParser

ДИСТРИБУТИВНАЯ МОДЕЛЬ ДЛЯ РАЗЛИЧЕНИЯ СЕМАНТИЧЕСКОГО СХОДСТВА И АССОЦИАЦИИ

Трофимов И. В. (itrofimov@gmail.com),
Сулейманова Е. А. (yes2helen@gmail.com)

Институт программных систем РАН,
Переславль-Залесский, Россия

1. Introduction

Measuring semantic relatedness of words or concepts plays an important role in the tasks of text categorization, search query expansion and many others. Of particular interest is a more specific case of semantic relatedness—semantic similarity, reflecting categorical commonality of terms (concepts). Semantic similarity has its special applications, for instance, in the construction of thesauri and ontologies. This article is devoted to the methods of distributional modeling that can tell semantically similar words from otherwise related cases. The models are designed for differentiating pairs of similar Russian-language nouns from those of thematically related ones, based on their syntactic context. This research complements the state of the art presented during RUSSE—the First Workshop on Russian Semantic Similarity Evaluation [25].

In 2015, the first RUSSE Workshop performed a systematic comparison and evaluation of different approaches to developing distributional semantic models aimed at revealing and measuring the degree of semantic similarity[1] of terms. Distributional models of semantics encode meanings of words as vectors in a highly dimensional space of context words. Similarity of word meanings is then measured as similarity of vectors. Such context vectors can be formed in a multitude of ways. The Workshop revealed that, for Russian, skip-gram [22] models currently perform the best, although other distributional approaches are only slightly behind:

- a classical DSM [30], where vectors are composed of most frequent Russian nouns, verbs, adjectives and adverbs;
- the GloVe model [27]; its application to Russian is described in [19];
- the CBOW [22]; experiments with the model are reported in [16].

Our research is an attempt to develop distributional models aimed at differentiating between two kinds of semantic relations:

- relations that are based on shared intrinsic features and common category membership (similarity);
- relations that stem from thematic, or situational, co-occurrence and are not supported by taxonomical commonality (associations).

Associations are given lower weights by our models. For context vector composition, we use a selective syntactic dependency approach: we only include the words that have a specific dependency relationship with the target word. Our measures are of the global type, as opposed to contextual ones [3], in that we do not use any context for meaning disambiguation. For evaluation, in the absence of Russian-language gold standards for testing the ability of the model to discriminate similarity from association, we compiled our own dataset RuSim1000.

---

1 We retain the term ‘similarity’ here, as it was used by the Workshop organizers; the right term would be ‘relatedness’ (terminological issues will be discussed in the next section).

2. Terminological issues

The notion of similarity is central to the domain of distributional semantics. In [31], the general idea behind the famous Distributional Hypothesis—a set of statements attributed to different authors[2]—is summed up as follows: "[...] there is a correlation between distributional similarity and meaning similarity, which allows us to utilize the former in order to estimate the latter".

Whereas the former—distributional similarity—can have a clearly defined mathematical sense, there is no common understanding of what the latter is.

The intuitive notion of similarity has proved very hard to define precisely. Even psychology, where similarity is one of the most central theoretical constructs, has not come up with a commonly agreed definition.

Similarity in its broad sense is very flexible. Many seem to agree that similarity of two things can only be judged with respect to some X. These 'respects' for similarity are determined by factors that are intrinsic to the comparison process [21]. As a consequence, two concepts (or two words) are not intrinsically similar or dissimilar [4]. The relative importance of common and different features depends on the task or context (solving odd-one-out puzzles is an illustrative example).

Before we address similarity issues from the prospective of distributional semantics, it is worth noting that, as evidenced by cognitive research [18], humans have distinct neural systems for two types of knowledge: feature-based taxonomic (categorical) knowledge and thematic knowledge—the "grouping of concepts by participation in the same scenario or event" [23].

In line with this distinction, two basic types of conceptual relations are distinguished. The first is feature-based taxonomic relatedness—this relation is commonly referred to as ***(semantic) similarity*** [10], [13], [1]. Semantic similarity refers to sharing of 'intrinsic' (perceptual or functional) features that account for membership in the same semantic category. *Car* and *bike* are said to be semantically similar because of their common physical features (wheels), their common function (transport), or because they fall within a clearly definable category (modes of transport) (example taken from [13]). Other terms for semantic similarity are *semantic category relatedness* [10], *taxonomic similarity / taxonomic relatedness* [17].

In contrast, the second type of relation—thematic relatedness—is based on co-occurrence (linguistic or otherwise) or functional relationships. Entities represented by thematically related concepts frequently occur together "in space and language" [13]. While similarity is based on feature overlap, thematically related concepts (entities) are not supposed to share intrinsic properties, although this is not impossible. In the psychological literature, thematic relatedness is mainly referred to as ***association*** [10], [13], although it is also known as *thematic similarity* [34], *topical similarity* [12], *domain similarity* [33], *thematic relatedness* [17], *relatedness* (as used in [1], but not in [7] and [8]—see below). Association is exemplified by pairs *car-petrol*, *bee-honey*.

---

2 The most cited version seems to be that of Rubenstein and Goodenough, 1965 (see reference item [29]): "words which are similar in meaning occur in similar contexts".

Similarity and association are two distinct relations, "neither mutually exclusive nor independent" [13]. Two related concepts can be

(1) similar and associated (*coffee-tea, brandy-wine, king-queen, doctor-nurse, blouse-skirt*);
(2) associated, but not similar (*coffee-mug, king-crown, engine-car, cow-milk*);
(3) similar, but not associated (*bear-cow, house-cabin, nurse-lawyer*).

2.1. Similarity versus association in NLP

With few exceptions [33], recent research in distributional semantics appears to have been focused on quantitative rather than qualitative aspects of word interaction within lexical semantic system. Such approaches neglect the difference between similarity and association [28], [14], [20], as their focus is estimating the strength of the connection between two words in the semantic network, regardless of the relation type. Such connection is most often referred to as ***relatedness*** [7], [8], [26] in the broad sense. Thus understood, semantic relatedness subsumes both semantic similarity and thematic association as its specific cases.

Until recently, there has been some confusion in terminology regarding the object of distributional semantic modeling within this paradigm. What is referred to as measuring 'similarity' as conveyed by distributional similarity turns out to be in fact estimating relatedness. Thus, the term 'semantic similarity' is sometimes taken to be the synonym for semantic relatedness and semantic proximity, and the inverse of semantic distance. For the sake of justice, it should be noted that in recent publications such terminological ambiguities are becoming rare.

Most of the gold standard datasets designed for the evaluation of distributional semantic models do not distinguish between taxonomic, feature-based similarity and thematic association (WordSim-353 [11], MEN Test Collection [6], RUSSE HJ [25]).

The utility of such resources to the development and application of distributional models is limited. This is still more so, because "many researchers appear unaware of what their evaluation resources actually measure" [13].

Recently developed resources, like SimLex999 [13], WS-Sim and WS-Rel subsets of WordSim353 [1], are expected to fill this gap. It is questionable, though, to what extent these resources can serve to actually measure the ability of models to reflect similarity as opposed to association. The developers of WS-Sim, for example, turned to human subjects to separate between similar and (otherwise) related cases, but left the original WordSim353 scores intact.

3. RuSim1000 dataset

The dataset was developed with the aim of evaluating Russian-language distributional models that focus on revealing *similarity* (possibly accompanied by association) as opposed to pure *association*.

Similarity, or taxonomical, feature-based similarity,—semantic relation that is based on shared intrinsic features and common category membership.

Association—semantic relation that stems from thematic, or situational, co-occurrence and is not supported by taxonomical (ontological) commonality.

RuSim1000 is composed of 1000 pairs of *related* nouns that are divided into two subsets—the sets of positive and negative examples. Positive examples are pairs of similar nouns. Negative examples are pairs of associated, but not similar nouns. Pairs of similar words that are also associated (*король-королева, king-queen*) are positive examples: it is the presence or absence of similarity that matters.

The core of the positive subset is formed by the following cases:
- synonyms (*имя-название, name-title*) and near synonyms (*особенность-аспект, peculiarity-aspect*);
- hyponym-hypernym (*имя-прозвище, name-nickname*) and the inverse (*питон-змея, python-snake*);
- co-hyponyms (*писатель-поэт, writer-poet*).

Clear-cut negative cases are pairs of nouns representing ontologically different entities linked by any of the following relations:
- part-whole (*шерсть-животное, fur-animal*) and the inverse (*лошадь-грива, horse-mane*);
- element-set (*самолет-эскадрилья, airplane-squadron*) and the inverse;
- functional (situational) relationship (*доктор-клиника, doctor-clinic, винтовка-выстрел, rifle-shot*);
- free association (*край-земля, edge-land*).

For a number of difficult and borderline cases the following decisions were made:
- Antonyms. Contrary to the intuition that antonymous terms are dissimilar, we take them to be similar (i.e. positive examples)—due to an assumption that their opposition is likely to hold within a certain category, to which they both belong (*свет-тьма, light-darkness*).
- Roles. As long as the taxonomy of roles should be separate from the taxonomy of types (or, at least, the conceptual difference between types and roles should be taken into account by the ontology), the category-membership criterion is somewhat difficult to apply to those pairs that are a mixture of a type and its role. It was decided to qualify as positive (i.e. similar):
    - pairs of the kind “a type and its typical role” (*торф-топливо, peat-fuel*, but not *самолет-вооружение, airplane-armament*);
    - thematically related roles of the same holder type, including complementary roles (*врач-медсестра, doctor-nurse, врач-пациент, doctor-patient*).

Table 1. Positive (similar words) and negative (associated, but not similar words) examples from RuSim1000

word1	word2	sim
лошадь (horse)	жеребец (stallion)	1
лошадь (horse)	кобыла (mare)	1
лошадь (horse)	пони (pony)	1
лошадь (horse)	кляча (jade)	1
лошадь (horse)	седло (saddle)	0
лошадь (horse)	конюх (groom)	0
лошадь (horse)	грива (mane)	0
лошадь (horse)	галоп (gallop)	0

RuSim1000 was designed in such a way that it would be compatible with the RUSSE evaluation framework[3]. Average Precision (AP) [35] used by RUSSE RT was chosen as evaluation measure. AP is calculated for a ranked list of examples. The higher the rank of the positive examples, the more they contribute to the AP (see formula 1).

$$Average\ Precision = \frac{\sum_r P@r}{R} \qquad (1)$$

where r is the rank of each positive example, R is the total number of positive examples, $P@r$ is the precision of the top-r examples.

Positive and negative examples in RuSim1000 are equal in number. As a consequence, the random baseline is about 0.5 [5]. It is carefully observed that there is equal number of positive and negative pairs beginning with the same word. Thus we could evaluate our algorithms with the same evaluation tools as were used for RUSSE RT.

The dataset is available at: https://zenodo.org/record/546238#.WPDyi6IlGUk

4. Semantic similarity measures

The way objects, events, phenomena etc. are categorized by humans is very much dependent on their activities. Despite rather flexible and 'non-systemic' character of categorization, which makes formal definition of shared categorical membership in terms of feature overlap almost impossible, concepts in many categories do share perceptual and functional features. The intuition behind our distributional models of similarity is as follows:

- similar objects tend to have more shared features than dissimilar;
- similar objects tend to act in similar way;
- similar objects tend to be exposed to similar actions.

That means we expect our similarity measures to have a clear interpretation as similarity of features and behavior.

[3] http://russe.nlpub.ru/downloads/

Put in linguistic terms, the three above statements (roughly) read as follows:

- similarity of features is evidenced as sharing common adjectives;
- similarity of behavior manifests itself as being the subject and/or the object of the same verbs.

Three types of syntactic relations (as defined in SynTagRus [2]) are retrieved from the source corpus:

- attributive (for feature-based similarity);
- predicative and 1-completive (for behavioral similarity).

The context vector is composed of adjectives, for feature-based similarity measure, and of verbs—for behavioral similarity. The length of vectors is not limited.

The idea of using syntactic features for distributional modeling is not new. There is a comprehensive survey on the topic in [24]. The authors suggest an elegant general framework for the integration of syntax-based and word co-occurrence approaches. A syntax-based approach to vector formation (for Russian word categorization) is presented in [15]. There is a hypothesis that models that learn from input annotated for syntactic or dependency relations better reflect similarity, whereas approaches that learn from running-text or bag-of-words input better model association [13]. However, we do not know of any attempt of applying syntax-based approaches for discriminating between similarity and association.

Back to our model, two questions are to be answered:

- how to form context vectors out of absolute frequencies of the target syntactic relations;
- how to measure the distance between vectors.

As far as English is concerned, Bullinaria and Levi [9] showed that the following combination proved to be working well for semantic relatedness evaluation: the positive pointwise mutual information (formula 2) as vector component value and cosine similarity for measuring the distance between vectors.

Let

t be the target noun,

c—a context word (a component of a context vector).

Then the pointwise mutual information *pmi(c, t)* is calculated as:

$$pmi(c,t) = log_2 \frac{p(c,t)}{p(c) \cdot p(t)},$$

where *p(c,t)* is the probability that *t* and *c* occur linked by the target relation,
p(c) and *p(t)* are the probabilities of independent occurrence of *c* and *t*, respectively.

The positive pointwise mutual information *ppmi(c, t)*:

$$ppmi(c,t) = \begin{cases} pmi(c,t), if\ pmi(c,t) \geq 0; \\ 0, \quad\quad if\ pmi(c,t) < 0. \end{cases} \tag{2}$$

We used ppmi with a single reservation: the probabilities *p(c)* and *p(t)* were calculated on the set of relations of the given type rather than on the entire corpus. To put it otherwise, for feature-based measure, for example, *p(c)* и *p(t)* are calculated on the set of noun-adjective pairs extracted from the corpus.

Besides, the cosine distance was taken to be zero if there were less than 10 non-zero summands in the numerator (formula 3).

$$\cos(A,B) = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\|A\| \cdot \|B\|} \quad (3)$$

5. Experiments and results

As a source of statistical information about syntactic relations we used the RuWac[4] corpus that had been syntactically annotated with MaltParser[5] [32]. A total of about 223 million target relation instances were extracted (Table 2).

Table 2. The statistical data obtained from RUWAC (the figures are given in million pairs)

	attributive	predicative	1-completive	total
number of relation instances	114	51	57	223
number of unique relations (lexeme$_1$, lexeme$_2$, relation_ type)	11.1	9.2	8.7	29

The data was used to develop distributional models that were evaluated against the RuSim1000 dataset.

The results of the tests[6] are presented in Tables 3 and 4. The Table 3 lists the average precision scores for the similarity/association discrimination task obtained by three models that learnt from single-syntactic-relation annotated input. Table 4 shows the average precision of the same categorization yielded by the models that learnt from combinations of two syntactic relations.

Table 3. Testing results for single-relation input

syntactic relation		
attributive	predicative	1-completive
0.907	0.846	0.882

Table 4. Testing results for combined input

combination of syntactic relations	
attributive + predicative	attributive + 1-completive
0.918	0.925

4 http://corpus.leeds.ac.uk/tools/ru/ruwac-parsed.out.xz (as of December 2016)

5 http://corpus.leeds.ac.uk/mocky/

6 The tests were performed with the software used for RUSSE—http://russe.nlpub.ru/downloads/

The experiments confirmed that a rather limited one- or two-relation syntactic context is sufficient to discriminate between similar and associated cases (this task being quite different from that of similarity or association scoring).

6. Conclusions

The paper focuses on the two cognitive and linguistic phenomena that account for semantic relatedness of terms—those of taxonomic similarity and thematic association. Dataset RuSim1000 is presented—a gold standard that can be used to evaluate the ability of models to discriminate between the two types of conceptual relations. Distributional models for similarity/association discrimination were developed, in which syntactic features of terms were used as 'proxies' for feature-based and behavioral similarity of objects. The experiments proved that the models are good enough at the task in hand (average 0.9 on RuSim1000).

The research was carried out as part of the project "Methods for automated extraction of events from texts" (No. AAAA-A17-117040610371-7).

References

1. *Agirre E., Alfonseca E., Hall K., Kravalova J., Pasca M., and Soroa A.* (2009), A study on similarity and relatedness using distributional and Wordnet-based approaches, Proceedings of NAACL, Boulder, pp. 19–27.
2. *Apresian Ju. D., Boguslavsky I. M., Iomdin B. L. et al.* (2005), Syntactically and Semantically Annotated Corpus of Russian: State-of-the-Art and Prospects [Sintaksicheski i semanticheski annotirovannyj korpus russkogo jazyka (sovremennoe sostojanie i perspektivy], National Corpus of Russian 2003–2005 (Results and Prospects) [Natsionalnyj korpus russkogo jazyka 2003–2005 g. (rezul'taty i perspektivy)], Moscow, Indrik, pp. 193–214.
3. *Arefyev N. V., Panchenko A. I., Lukanin A. V., Lesota O. O., Romanov P. V.* (2015), Evaluating three corpus-based semantic similarity systems for russian, Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", RGGU, pp. 106–118.
4. *Batchkarov M., Kober T., Reffin J., Weeds J., Weir D.* (2016), A critique of word similarity as a method for evaluating distributional semantic models, Proceedings of The First Workshop on Evaluating Vector Space Representations for NLP, Berlin, pp. 7–12.
5. *Bestgen Y.* (2015), Exact Expected Average Precision of the Random Baseline for System Evaluation, The Prague Bulletin of Mathematical Linguistics, 103, pp. 131–138.
6. *Bruni E., Tran N. Kh., Baroni M.* (2014), Multimodal Distributional Semantics, Journal of Artificial Intelligence Research (JAIR), 49, pp. 1–47.
7. *Budanitsky A., Hirst G.* (2001), Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures, Workshop on WordNet and Other Lexical Resources, Second Meeting of the North American Chapter of the Association for Computational Linguistics, Pittsburgh, pp. 29–34.

8. *Budanitsky A., Hirst G.* (2006), Evaluating Wordnet-based measures of lexical semantic relatedness, Computational Linguistics, 32(1), pp. 13–47.
9. *Bullinaria J., Levy J.* (2007), Extracting Semantic Representations from Word Co-occurrence Statistics: A Computational Study, Behavior Research Methods, 39(3), pp. 510–526.
10. *Chiarello C., Burgess C., Richards L., Pollock A.* (1990), Semantic and associative priming in the cerebral hemispheres: some words do, some words don't ... sometimes, some places, Brain and Language, 38, pp. 75–104.
11. *Finkelstein L., Gabrilovich E., Matias Y., Rivlin E., Solan Z., Wolfman G., Ruppin E.* (2001), Placing search in context: The concept revisited, Proceedings of the Tenth International World Wide Web Conference, Hong Kong, pp. 406–414.
12. *Hatzivassiloglou V., Klavans J. L., Holcombe M. L., Barzilay R., Kan M.-Y., McKeown K.* (2001), Simfinder: A flexible clustering tool for summarization, Proceedings of the NAACL Workshop on Automatic Summarization, Pittsburgh, available at: http://hdl.handle.net/10022/AC:P:20214.
13. *Hill F., Reichart R., Korhonen A.* (2015), SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation, Computational Linguistics 41(4), pp. 665–695.
14. *Huang E. H., Socher R., Manning C. D., Ng A. Y.* (2012), Improving word representations via global context and multiple word prototypes, Proceedings of ACL, Jeju Island, Korea, pp. 873–882.
15. *Klyshinsky E. S., Kochetkova N. A., Logacheva V. K.* (2013), Clustering words with similar sense using information about their syntactic dependencies [Metod klasterizacii slov s ispol'zovaniem informacii ob ikh sintaksicheskoj sviaznosti], Scientific and technical information. Series 2: Information processes and systems [Nauchno-tekhnicheskaja informacija. Serija 2: Informacionnye processy i sistemy], 11, pp. 36–43.
16. *Kutuzov A., Andreev I.* (2015), Texts in, Meaning out: Neural Language Models in Semantic Similarity Tasks for Russian, Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", RGGU, pp. 133–144.
17. *Landrigan J.-F., Mirman D.* (2016), Taxonomic and Thematic Relatedness Ratings for 659 Word Pairs, Journal of Open Psychology Data, 4: e2.
18. *Lewis G. A., Poeppel D., Murphy G. L.* (2015), The neural bases of taxonomic and thematic conceptual relations: An MEG study, Neuropsychologia, 68, pp. 176–189.
19. *Lopukhin K. A., Lopukhina A. A., Nosyrev G. V.* (2015), The Impact of Different Vector Space Models and Supplementary Techniques on Russian Semantic Similarity, Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", RGGU, pp. 145–153.
20. *Luong M.-Th., Socher R., Manning C. D.* (2013), Better word representations with recursive neural networks for morphology, Proceedings of the Seventeenth Conference on Computational Natural Language Learning (CoNLL-2013), Sofia, pp. 104–113.

21. *Medin D. L., Goldstone R. L., Gentner D.* (1993), Respects for similarity, Psychological Review, 100(2), pp. 254–278.
22. *Mikolov T., Chen K., Corrado G., Dean J.* (2013), Efficient Estimation of Word Representations in Vector Space, arXiv preprint arXiv:1301.3781.
23. *Mirman D., Walker G. M., Graziano K. M.* (2011), A Tale of Two Semantic Systems: Taxonomic and Thematic Knowledge, Proceedings of the 33th Annual Meeting of the Cognitive Science Society (CogSci 2011), Boston, pp. 2211–2216.
24. *Pado S., Lapata M.* (2007), Dependency-based Construction of Semantic Space Models, Computational Linguistics, 33(2), pp. 161–199.
25. *Panchenko A., Loukachevitch N., Ustalov D., Paperno D., Meyer C. M., Konstantinova N.* (2015), RUSSE: The First Workshop on Russian Semantic Similarity, Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", RGGU, pp. 89–105.
26. *Panchenko A., Ustalov D., Arefyev N., Paperno D., Konstantinova N., Loukachevitch N. and Biemann C.* (2016), Human and Machine Judgements about Russian Semantic Relatedness, Proceedings of the 5th Conference on Analysis of Images, Social Networks, and Texts (AIST'2016), Communications in Computer and Information Science (CCIS), Springer-Verlag Berlin Heidelberg, pp. 221–235.
27. *Pennington J., Socher R., Manning C. D.* (2014), Glove: Global vectors for word representation, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543.
28. *Reisinger J., Mooney R. J.* (2010), A mixture model with sharing for lexical semantics, Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP), Massachusetts, pp. 1173–1182.
29. *Rubenstein H., Goodenough J. B.* (1965), Contextual correlates of synonymy, Communications of the ACM (Comp. linguistics), 8(10), pp. 627–633.
30. *Ryzhova D. A., Kyuseva M. V.* (2015), On the Nature of Semantic Similarity and It's Measuring with Distributional Semantics Models, available at: http://www.dialog-21.ru/digests/dialog2015/materials/pdf/RyzhovaDAKyusevaMV.pdf.
31. *Sahlgren M.* (2008), The Distributional Hypothesis, Rivista di Linguistica (Italian Journal of Linguistics), 20 (1), pp. 33–53.
32. *Sharoff S., Nivre J.* (2011), The proper place of men and machines in language technology: Processing Russian without any linguistic knowledge, Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", RGGU, pp. 591–604.
33. *Turney P. D.* (2012), Domain and Function: A Dual-Space Model of Semantic Relations and Compositions, Journal of Artificial Intelligence Research (JAIR), 44, pp. 533–585.
34. *Wisniewski E. J., Bassok M.* (1996), On putting milk in coffee: the effect of thematic relations on similarity judgments, Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society, Erlbaum, pp. 464–468.
35. *Zhang E., Zhang Y.* (2009), Average Precision, Encyclopedia of Database Systems, Springer US, pp. 192–193.