

Назва	Building the Semantic Similarity Model for Social Network Data Streams
Автори	Svitlana Petrasova, Nina Khairova, Wlodzimierz Lewoniewski
Ключові слова	social network, data stream, collocations, semantic similarity, blogs, corpus, Universal Dependencies, WordNet
Дата публікації	2018
Видавництво	IEEE
Бібліографічний опис	Petrasova, S., Khairova, N., Lewoniewski, W.: Building the Semantic Similarity Model for Social Network Data Streams. In: 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), pp. 21–24, Lviv (2018).
DOI	10.1109/DSMP.2018.8478480
Реферат	This paper proposes the model for searching similar collocations in English texts in order to determine semantically connected text fragments for social network data streams analysis. The logical-linguistic model uses semantic and grammatical features of words to obtain a sequence of semantically related to each other text fragments from different actors of a social network. In order to implement the model, we leverage Universal Dependencies parser and Natural Language Toolkit with the lexical database WordNet. Based on the Blog Authorship Corpus, the experiment achieves over 0.92 precision.
References	<ol style="list-style-type: none"> 1. M. Adedoyin-Olowe, M. M. Gaber and S. Frederic, "A Survey of Data Mining Techniques for Social Media Analysis", Journal of Data Mining & Digital Humanities, 2014. 2. J. Golbeck, Analyzing the Social Web. M. Kaufmann, 2013. 3. J. Scott and P.J. Carrington, The SAGE Handbook of Social Network Analysis. SAGE Publications, 2011. 4. D. M. Boyd and N. B. Ellison, "Social Network Sites: Definition History and Scholarship", Journal of Computer-Mediated Communication, vol. 13, no. 1, pp. 210-230, 2007. 5. M. J. F. Rodrigues and A. J. S. Teixeira, Advanced Applications of Natural Language Processing for

Performing Information Extraction. Springer, 2015.

6. A. Pesaranhader, A. Pesaranhader, S. Matwin and M. Sokolova, "One Single Deep Bidirectional LSTM Network for Word Sense Disambiguation of Text Data", Springer Advances in Artificial Intelligence: 31 st Canadian Conference on Artificial Intelligence Canada, pp. 96-107, 2018.

7. Hua Wu and Ming Zhou, "Synonymous Collocation Extraction Using Translation Information", 41st Annual Meeting on Association for Computational Linguistics (ACL'03) Stroudsburg PA USA, vol. 1, pp. 120-127, 2003.

8. M. Pasca and P. Dienes, "Aligning Needles in a Haystack: Paraphrase Acquisition Across the Web", Second International Joint Conference: Natural Language Processing (IJCNLP 2005) Korea, pp. 119-130, 2005.

9. "Extracting paraphrases from a Parallel Corpus", 39th Annual Meeting on Association for Computational Linguistics (ACL '01), pp. 50-57, 2001.

10. S. Petrasova and N. Khairova, "Automatic Identification of Collocation Similarity", 10th International Scientific and Technical Conference: Computer Science & Information Technologies (CSIT'2015) Lviv, pp. 136-138, 2015.

11. S. Petrasova and N. Khairova, "Using a Technology for Identification of Semantically Connected Text Elements to Determine a Common Information Space" in Cybernetics and Systems Analysis Springer, vol. 53, no. 1, pp. 115-124, 2017.

12. M. Bondarenko and Yu. Shabanov-Kushnarenko, The intelligence theory. Kharkiv SMIT, 2007.

13. P. Basile, A. Caputo and G. Semeraro, "An Enhanced Lesk Word Sense Disambiguation Algorithm through a Distributional Semantic Model", International Conference on Computational Linguistics, pp. 1591-1600, 2014.

14. J. Schler, M. Koppel, S. Argamon and J. Pennebaker, "Effects of Age and Gender on Blogging", 2006 AAI Spring Symposium on Computational Approaches for Analyzing Weblogs, pp. 191-197, 2006.

15. T. Pedersen, S. Patwardhan and J. Michelizzi, "WordNet:: Similarity-Measuring the Relatedness of Concepts", Demonstration Papers at HLT-NAACL, pp. 38-41, 2004.

16. W. Lewoniewski, "Enrichment of Information in Multilingual Wikipedia Based on Quality Analysis", International Conference on Business Information Systems

	Springer Cham, pp. 216-227, 2017.
Location	https://ieeexplore.ieee.org/document/8478480