Назва	Open Information Extraction as Additional Source for Kazakh Ontology Generation
Автори	Nina Khairova, Svitlana Petrasova, Orken Mamyrbayev, Kuralay Mukhsina
Ключові слова	open information extraction, RDF-triplets, unstructured text, logical-linguistic equations, Kazakh bilingual news websites
Дата публікації	2020
Видавництво	Springer
Бібліографічний опис	Khairova, N., Petrasova, S., Mamyrbayev, O., Mukhsina, K.: Open Information Extraction as Additional Source for Kazakh Ontology Generation. In: Nguyen N., Jearanaitanakij K., Selamat A., Trawiński B., Chittayasothorn S. (eds) Intelligent Information and Database Systems. ACIIDS 2020. Lecture Notes in Computer Science 12033, pp. 86–96, Switzerland (2020).
DOI	10.1007/978-3-030-41964-6_8
Реферат	Nowadays, structured information that obtains from unstructured texts and Web context can be applied as an additional source of knowledge to create ontologies. In order to extract information from a text and represent it in the RDF-triplets format, we suggest using the Open Information Extraction model. Then we consider the adaptation of the model to fact extraction from unstructured texts in the Kazakh language. In our approach, we identify lexical units that name the participants of the action (the Subject and Object) and semantic relations between them based on words characteristics in a sentence. The model provides semantic functions of the action participants via logical-linguistic equations that express the relations of the grammatical and semantic characteristics of the words in a Kazakh sentence. Using the tag names and some syntactic characteristics of words in the Kazakh sentences as the values of the predicate variables in corresponding equations allows us to extract Subjects, Objects and Predicates of facts from texts of Web content. The experimental research dataset includes texts extracted from Kazakh bilingual news websites. The experiment shows that we can achieve the precision of facts extraction over 71% for Kazakh corpus.
References	1. Sint, R., Schaffert, S., Stroka, S., Ferstl, R.: Combining

unstructured, fully structured and semi-structured information in semantic wikis. In: Proceedings of the 4th Semantic Wiki WorkShop (SemWiki) at the 6th European Semantic Web Conference, ESWC (2009)
2. Crestan, E., Pantel, P.: Web-scale knowledge extraction from semi-structured tables. In: WWW 2010 Proceedings of the 19th International Conference on World Wide Web, pp. 1081–1082 (2010)
3. Wong, Y.W., Widdows, D., Lokovic, T., Nigam, K.: Scalable attribute-value extraction from semi-structured text. In: 2009 IEEE International Conference on Data Mining Workshops, pp. 302–307 (2009)
4. Phillips, W., Riloff, E.: Exploiting strong syntactic heuristics and co-training to learn semantic lexicons. In: Proceedings of the conference on Empirical Methods in Natural Language Processing (EMNLP) (2002)
5. Jones, R., Ghani, R., Mitchell, T., Riloff, E.: Active learning with multiple view feature sets. In: ECML 2003 Workshop on Adaptive Text Extraction and Mining (2003)
6. ARPA. Proceedings of the 3rd Message Undestanding Conference (1991)
7. Etzioni, O., Banko, M., Soderland, S., Weld, D.: Open information extraction from the web. Commun. ACM 51(12), 68–74 (2008)
8. Fader, A., Soderland, S., Etzioni, O.: Identifying relations for open information extraction. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK, pp. 1535–1545 (2011)
9. Duc-Thuan, V., Ebrahim, B.: Open information extraction. In: Encyclopedia with Semantic Computing and Robotic intelligence, vol. 1, no. 1 (2016)
10. Shinzato, K., Sekine, S.: Unsupervised extraction of attributes and their values from product description. In: Sixth International Joint Conference on Natural Language Processing, IJCNLP 2013, pp. 1339–1347 (2013)
11. Liu, L., Ren, X., Zhu, Q., et al.: Heterogeneous supervision for relation extraction: a representation learning approach. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 46–56 (2017)
12. Wang, X., Zhang, Y., Chen, Y.: A survey of truth discovery in information extraction (2018)
13. Gamallo, P., Garcia, M., Fernandez-Lanza, S.: Dependency-based open information extraction. In:

Proceedings of the Joint Workshop on Unsupervised and Semi-Supervised Learning in NLP, pp. 10–18 (2012)
14. Akbik, A., Loser, A.: KrakeN: N-ary facts in open information extraction. In: Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction, pp. 52–56 (2012)
15. Fader, A., Soderland, S., Etzioni, O.: Identifying relations for open information extraction. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1535–1545 (2011)
16. Angeli, G., Premkumar, M.J., Manning, C.D.: Leveraging linguistic structure for open domain information extraction. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, pp. 344–354 (2015)
17. Gashteovsk, K., Gemulla, R., Del Corro, L.: MinIE: minimizing facts in open information extraction. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 2630–2640 (2017)
18. Mooney, R.J., Bunescu, R.: Mining knowledge from text using information extraction. ACM SIGKDD Explor. Newslett. 7(1), 3–10 (2005). Natural language processing and text mining
19. Gamallo, P., Garcia, M.: Multilingual open information extraction. In: Portuguese Conference on Artificial Intelligence, pp. 711–722 (2015)
20. Khairova, N., Lewoniewski, W., Węcel, K.: Estimating the quality of articles in russian wikipedia using the logical-linguistic model of fact extraction. In: Abramowicz, W. (ed.) BIS 2017. LNBIP, vol. 288, pp. 28–40. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59336-4_3
21. Khairova, N., Lewoniewski, W., Węcel, K., Orken, M., Kuralai, M.: Comparative analysis of the informativeness and encyclopedic style of the popular web information sources. In: Abramowicz, W., Paschke, A. (eds.) BIS 2018. LNBIP, vol. 320, pp. 333–344. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93931-5_24
22. Khudhair, A.T.: The intelligence theory mathematical apparatus formal BASE. Adv. Inf. Syst. 1(1), 38–43 (2017)
23. Khairova, N.F., Petrasova, S., Gautam, A.P.S.: The logical-linguistic model of fact extraction from English texts. In: Dregvaite, G., Damasevicius, R. (eds.) ICIST 2016. CCIS, vol. 639, pp. 625–635. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46254-7_51

	24. Regneri, M., Wang, R.: Using discourse information for paraphrase extraction. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 916–927 (2012)
Location	https://link.springer.com/chapter/10.1007/978-3-030-41964-6_8