

Назва	The Logical-Linguistic Model of Fact Extraction from English Texts
Автори	Nina Feliksivna Khairova, Svetlana Petrasova, Ajit Pratap Singh Gautam
Ключові слова	logical-linguistic model, fact extraction, finite predicates algebra, triplet, grammatical features, syntactic features
Дата публікації	2016
Видавництво	Springer
Бібліографічний опис	Khairova, N., Petrasova, S., Gautam, A. P. S.: The Logical-Linguistic Model of Fact Extraction from English Texts. In: Dregvaite G., Damasevicius R. (eds) Information and Software Technologies. ICIST 2016. Communications in Computer and Information Science 639, pp. 625–635, Switzerland (2016).
DOI	10.1007/978-3-319-46254-7_51
Реферат	In this paper we suggest the logical-linguistic model that allows extracting required facts from English sentences. We consider the fact in the form of a triplet: Subject > Predicate > Object with the Predicate representing relations and the Object and Subject pointing out two entities. The logical-linguistic model is based on the use of the grammatical and semantic features of words in English sentences. Basic mathematical characteristic of our model is logical-algebraic equations of the finite predicates algebra. The model was successfully implemented in the system that extracts and identifies some facts from Web-content of a semi-structured and non-structured English text.
References	<ol style="list-style-type: none"> 1. Fader, S., Soderland, O.: Etzioni Identifying relations for open information extraction. In: Conference on Empirical Methods in Natural Language Processing. Edinburgh, Scotland, pp. 1535–1545 (2011) 2. 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) 3. 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)

4. Gatterbauer, W., Bohunsky, P., Herzog, M., Krupl, B., Pollak, B.: Towards domain-independent information extraction from web tables. In: Proceedings WWW-07, pp. 71–80. Banff, Canada (2007)
5. 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)
6. 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)
7. 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)
8. Agichtein, E., Gravano, L.: Snowball: extracting relations from large plaintext collections. In: Proceedings of the 5th ACM International Conference on Digital Libraries, pp. 85–94. San Antonio, Texas (2000)
9. Ludovic, L., Gallinari, P.: Bayesian network model for semi-structured document classification. *Inf. Proc. Manage. Int. J. Spec. Issue Bayesian Netw. Inf. Retrieval* 40, 807–827 (2004)
10. Rish, I.: An empirical study of the naive bayes classifier. In: Proceedings of IJCAI-01 Workshop on Empirical Methods in Artificial Intelligence (2001)
11. Jatana, N., Sharma, K.: Bayesian spam classification: time efficient radix encoded fragmented database approach. In: 2014 International Conference on Computing for Sustainable Global Development (INDIACom), pp. 939–942 (2014)
12. Aiwu, L., Hongying, L.: Utilizing improved bayesian algorithm to identify blog comment spam. In: IEEE Symposium on Robotics and Applications (ISRA), pp. 423–426 (2012)
13. Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In: ECML 1998 Proceedings of the 10th European Conference on Machine Learning, pp. 137–142. Springer-Verlag London, UK (1998)
14. Kleinbaum, D.G., Klein, M., Pryor, E.R.: Logistic Regression: A Self-Learning Text. Springer, New York (2002)
15. Baoli, L., Shiwen, Y., Qin, L.: An improved k-nearest neighbor algorithm for text categorization. In: The 20th

	<p>International Conference on Computer Processing of Oriental Languages, Shenyang, China (2003)</p> <p>16. Manne, S., Kotha, S. K., Fatima, S.: Text Categorization with k-nearest neighbor approach. In: Proceedings of the International Conference on Information Systems Design and Intelligent Applications, vol.132, pp. 413–420 (2012)</p> <p>17. Entezari-Maleki, R., Rezaei, A., Minaei-Bidgoli, B.: Comparison of classification methods based on the type of attributes and sample size. <i>J. Convergence Inf. Technol. (JCIT)</i> 4(3), 94–102 (2009)</p> <p>18. Mooney, R.J., Bunescu, R.: Mining knowledge from text using information extraction. <i>Newsl. ACM SIGKDD Explor. Newsl. Nat. Lang. Process. Text Min.</i> 7(1), 3–10 (2005)</p> <p>19. Yahya, M., Whang, E.S., Gupta R., Halevy A.: ReNoun: fact extraction for nominal attributes. In: Proceedings of the Conference on Empirical Methods in Natural Language (EMNLP), pp. 325–335 (2014)</p> <p>20. Luckicgeev, S.: Graphical notations for rule modeling. In: Giurca, A., Gašević, D., Taveter, K. (eds.) <i>Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches</i>, Hershey, New York., vol. 1, pp. 76–98 (2009)</p> <p>21. Bondarenko, M.: Shabanov-Kushnarenko, J. 2007. The intelligence theory. Kharkiv: “SMIT”, 576. (In Russian)</p> <p>22. Khairova, N., Sharonova, N., Gautam, A.P.: Logic-linguistic model of fact generation from text streams of corporate information system. <i>Int. J. Inf. Theor. Appl.</i> 22(2), 142–152 (2015)</p>
Location	https://link.springer.com/chapter/10.1007/978-3-319-46254-7_51