Title: "Modeling an Intelligent System for Extracting Information of a Certain Type from Text"

Author: Diana Agrba

Research Supervisor: Ph.D. in Economics Olga Timchenko

Research Initiator: OOO «Advertising Agency KAKADU»

Topical Importance: Currently, there is an extremely rapid development of Internet technologies, the Internet, where knowledge and information in large numbers are mostly in text form and the amount of unstructured information is only increasing, while currently less than 1% of texts are analyzed. The market growth is mainly for account analysis of social networking data, and therefore a particularly urgent task was the automated extraction of information from the text. Information can be made more structured by transforming it into a relational form or by adding XML markup. When monitoring news feeds with the help of intelligent agents, you just need methods of extracting information and converting it into a form that will be more convenient to work with later. Information extracted from the text can act as a meta-information, presenting a text document when solving information search, classification, clustering, annotation and abstracting problems. To solve this problem, an intelligent system for extracting information from the text in natural language will be modeled in this paper.

Goal: The purpose of this work is to study the process of extracting information from the text, its organization and modeling an intelligent system for extracting information from the text. The result of this work should be the implementation of the simulated system.

Tasks: The tasks of this work include the collection, systematization, processing of the actual material on the topic of the final qualifying work, the study of intellectual systems, the study of the problematics of extracting information from the text. Then the study of machine learning, the development of the model of the intellectual system, the consideration of the ways of implementing the created model and the implementation of an intelligent system for extracting information from text.

Theoretical and practical value: The theoretical significance of the obtained results is determined by the definition of the problem of extracting information (in particular keywords) from the text, by considering and identifying ways to solve this problem and tasks, systematizing available information on the problem area, methods developed and implementing their algorithms. Also justifying the choice of directions for future research, which eventually led to the creation of an intelligent system for extracting information from the text, which makes it easier and faster to search information and knowledge on the Internet and provides a practical importance of this final qualifying work.

Results: A large amount of information and material was studied, goals were achieved and the tasks were accomplished. An intelligent system for extracting information of a certain type from the text was modeled, and then implemented using a high-level programming language Python. As the extracted information,
the keywords were chosen, which allow also to extract information about the topic of the text, using the methods of machine learning. After training and testing on more than three thousand markup texts, the result was obtained in 89% of the accuracy of the developed program-classifier (intellectual system).

**Implementation advice:** To obtain better, more accurate results of the developed intellectual system, the database of training texts should be replenished with new materials. Due to the fact that the intellectual system still inaccurately determines the theme of the text, if it is not enough or too voluminous, it is necessary to refine its algorithm and choose the best classification parameters.