

ABSTRACT

The bachelor's thesis: 63 p., 17 fig., 8 tabl., 2 appendices and 15 sources.

The theme of this thesis is “Keywords Queries Generation for Automation SEO-specialists work”.

The object of study – sites and methods of promotion in the search engines.

Subject of research – methods for classifying text documents for filtering search queries based on keywords by type.

The aim of this work is to develop a module of keywords queries generation and a user interface for it.

Research methods – logistic regression, passive-aggressive classifier, AdaBoost.

Relevance – available tools for query generation do not provide possibility of their filtering by information and commercial types. The distinction of these types of queries is important in the formation of the semantic kernel.

Performance results – an algorithm for search query classification; the proposed configuration of classifier was implemented; an application with graphical user interface to demonstrate the module's work was developed.

Further development of the research subject – improve data preparation by using named-entity recognition.

CLASSIFICATION, MACHINE LEARNING, NATURAL LANGUAGE PROCESSING, PORTER STEMMER, TF-IDF, LOGISTIC REGRESSION, ADABOOST, PASSIVE-AGGRESSIVE CLASSIFIER.