

ANNOTATION

Theme of the final qualifying work: "Development of an intelligent system of automatic text abstracting (based on Russian texts)"

Author: Burkadze Nana Gochaevna

Scientific leader of the WRC: Cand. econ. of Sciences Timchenko Olga Viktorovna

Information about the contracting authority: LLC "Advertising agency CACADE"

The relevance of the research topic: Modern development of society is the most expensive resource is time. A person is constantly confronted with a lot of information of a different kind and representation, and despite the abundance and variety of information, it is forced to process it in a timely manner. Most of this information is represented by texts in natural language. In the case when there are too many documents and the person is not able to read them carefully in the time allotted for this purpose, the systems of automatic abstracting of text documents come to the rescue. Automatic abstracting and annotation of texts is one of the most important trends in the development of modern information technologies. The amount of information that a person has to deal with is constantly growing and the time comes when it becomes impossible to work out all the necessary material. The development of new algorithms for automatic abstracting and annotation of texts in natural language is still not lost its relevance, and is becoming increasingly necessary in connection with the ever-increasing volume of text data.

Objective: The aim of this work is to study the task of automatic text abstracting and to find new approaches to solving this problem using technologies based on intelligent systems. The projected system of automatic text abstracting should improve the semantic quality of the abstract and improve the efficiency of data processing and knowledge processes in computer systems, and also allow the processing of texts of different genres, different levels of complexity of terminology and volumes.

Objectives: The main tasks of the research are the review of the current state of the problem of automatic text abstracting, the study of text abstracting methods, the review of existing text abstracting systems, the development of the architecture of the projected system, the development of an algorithm for automatic text abstracting, the development of a software product that implements the developed algorithm for automatic text abstracting, The proposed algorithm and known algorithms for solving the given th problem. As a result of the work, the structure of the automatic text summarizing system should be developed, methods selected and algorithms developed that need to be implemented in its modules, as well as ways to improve the quality of the developed system.

Theoretical and practical significance of the research: Theoretical and practical significance of the research consists in approbation of the algorithm of automatic abstracting of textual information proposed within the framework of the final qualifying work.

Results of the research: Within the framework of the final qualification work, the goal was achieved: an intelligent system of automatic text abstracting in Russian was designed. The system used statistical methods for preprocessing texts, such as the stamping algorithm based on the Mystem algorithm, the TF * IDF algorithm for determining keywords, the algorithm for determining the weighting characteristics vector (with the number of selected features taken equal to five). Statistical methods of text processing have generated input data for using fuzzy logic methods. A fuzzy centroid was used as a membership function. The rules of fuzzy inference are formulated in the form of "if-then". The algorithm used for fuzzy inference allowed for all sentences of the text to get the values of the membership function of the set "an important sentence". Using the quick sort algorithm, the sentences of the referenced text are sorted by descending the values of the membership function. The compression ratio specified by the user determines the number of sentences that must be included in the abstract. The efficiency of the intelligent automatic text summarizing system was evaluated.

Recommendations: To increase the efficiency of the system as a whole, it is necessary to continue training, to include the procedure for cleaning the text from stop words, service words, to the pre-processing stage. In addition, in the procedure of fuzzy inference, it is necessary to add "if-then" rules to take into account the linguistic rules of natural language.