

ABSTRACT

Relevance: the task of defining proper names in the text is an intermediate stage for many other tasks. The result of her work is to determine the relationship between proper names and the categories to which they belong. Further, these relationships are widely used to automate the construction of hierarchies and categories of unstructured documents, such as news articles. They are also used to implement tag search, which avoids full-text search to save computing resources and time. Another topical application is the implementation of sentimental analysis, when among the input data of any structure, such as data from social networks, proper names are determined, and the emotional color of the text that touches them.

Another topical issue is the extremely little attention paid to the tasks of natural language processing in the context of the Ukrainian language, and the task of defining proper names is no exception.

The purpose: to create a model for determining proper names in the text based on machine learning.

To achieve this goal, the following tasks were formulated:

- systematization of existing approaches to determining proper names;
- preparation of data for training;
- development of a model based on machine learning;
- training and testing of the model;
- analysis of the efficiency of the obtained algorithm;
- search for further research.

Object of research: definition of proper names in the text in Ukrainian.

Subject of research: algorithms for determining proper names in the text.

Research methods: neural networks, networks of long short-term memory, the method of random conditional fields were used to solve the problem.

Scientific novelty: the most important scientific results of the master's dissertation are research of possibilities of definition of own names in texts in the Ukrainian language at rather small sizes of training data; study of the dependences of the influence of the number of input data on the results of determining proper names in texts in Ukrainian and English; research of influence of normalization of text data on results of training of models of definition of proper names.

The practical value of the obtained results is determined by the fact that the proposed architecture of the model allows to achieve the value of average harmonic accuracy and completeness of 92% for texts in English and 72% for texts in Ukrainian, which allows to use it to determine proper names in arbitrary texts.

Relationship with working with scientific programs, plans, topics: the work was performed at the Department of Automated Information Processing and Control Systems of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute".

Approbation: the main provisions of the work were reported and discussed at the VI All-Ukrainian scientific-practical conference of young scientists and students "Information systems and management technologies" (ISTU-2021).

Keywords recognition of proper names, neural networks, machine learning, long-term short-term memory network, conditional random fields.