

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

Кафедра дифференциальных уравнений и системного анализа

Аннотация к дипломной работе

**ПРИМЕНЕНИЕ ВЕКТОРНОЙ МОДЕЛИ ПРЕДСТАВЛЕНИЯ
СУЩНОСТЕЙ ПРИ ЗАДАНИИ СВЯЗЕЙ В ОНТОЛОГИИ
МАТЕМАТИКИ**

Тишуров Алексей Петрович

Научный руководитель:
кандидат физ.-мат. наук,
доцент Л.Л. Голубева

2016

РЕФЕРАТ

Дипломная работа содержит 43 страницы, 6 рисунков, 5 таблиц, 2 приложения, 1 DVD-диск, 22 источника.

Ключевые слова: ВЕКТОРНЫЕ МОДЕЛИ, КЛАССИФИКАЦИЯ, КОРПУС ДОКУМЕНТОВ, ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА, PYTHON.

Дипломная работа представляет собой описание создания корпуса математических документов, его обработки и обучения векторных моделей семантики.

Целью дипломной работы является задание меры подобия математических статей на основе обученных моделей.

Для достижения поставленной цели использовались:

- математические статьи, находящиеся в открытом доступе;
- методы обработки естественного языка и векторные модели семантики word2vec и GloVe,
- методы машинного обучения в задаче классификации;
- алгоритм t -SNE визуализации многомерных данных;
- язык программирования Python.

В дипломной работе представлены следующие результаты:

- составлен уникальный корпус математических статей;
- предложен собственный метод выделения из математических документов максимального объема информации;
- предложен алгоритм выделения математических утверждений из текстов;
- описан математический аппарат двух методов построения векторных моделей семантики;
- обучены и протестированы векторные модели;
- задана мера подобия между статьями.

Дипломная работа выполнена автором самостоятельно.

ABSTRACT

The thesis is presented in form of explanatory note of 43 pages, 6 figures, 5 tables, 2 appedices, 1 DVD, 22 references.

Keywords: VECTOR MODELS, CLASSIFICATION, DOCUMENT CORPUS, NATURAL LANGUAGE PROCESSING, PYTHON.

The thesis project describes creation of mathematical document corpus, processing and learning vector models of semantics.

The goal of the thesis project is creation of articles similarity measure based on vector models.

To achieve the goal we used:

- mathematical articles that are in the public doamin;
- natiral language processing and vector model of semantics word2vec and GloVe;
- machinge learning methods in classification task
- τ -SNE algorithm for multidimensional data visualization;
- python programming language.

The main results of the thesis project are as follows:

- a unique corpus of mathematical articles;
- own method of the maximal amount of information extraction from mathematical articles;
- the algorithm of mathematical statements extaction from the text;
- the mathematical foundations of the two methods for costruction vector models were described;
- vector models were trained and tested;
- article similarity measure was proposed.

The thesis was written solely by the author.