

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра дифференциальных уравнений и системного анализа

Аннотация к магистерской диссертации

**ПРЕДСКАЗАНИЕ КОДОВ БОЛЕЗНЕЙ
ПО КЛИНИЧЕСКИМ ЗАМЕТКАМ**

Кулеш Антон Сергеевич

Научный руководитель:
кандидат физ.-мат. наук,
доцент А. Э. Малевич

2018

В магистерской диссертации 58 страниц, 23 рисунка, 8 таблиц, 30 источников, 1 приложение.

ВЕКТОРНАЯ МОДЕЛЬ, ДИСТРИБУТИВНАЯ СЕМАНТИКА, ЭЛЕКТРОННАЯ МЕДИЦИНСКАЯ КАРТА, МКБ-9, МНОГОМЕТКОВАЯ КЛАССИФИКАЦИЯ, РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ, СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ

В данной магистерской диссертации исследуется проблема автоматического тегирования электронных медицинских документов. Решается задача предсказания множества меток (кодов болезней по МКБ-9) по отчетам о выписке пациентов.

Целью работы является формализация проблемы автоматического тегирования документов, исследование и сравнение различных подходов для ее решения.

Результаты магистерской диссертации:

1. Сформирован корпус документов (около 53 000 случаев госпитализации)
2. Построены векторные модели корпуса документов, проведено сравнение различных подходов
3. Построены и проанализированы модели многометковой классификации, проведено сравнение качества и сложности моделей для наборов с различным количеством меток (50 и 1612)

Эксперименты реализованы на языке программирования Python.

Магистерская диссертация выполнена автором самостоятельно.

The master thesis is presented in form of explanatory note of 58 pages, 23 figures, 8 tables, 1 appendix, 30 references.

VECTOR SPACE MODEL, DISTRIBUTIONAL SEMANTICS, ELECTRONIC HEALTH RECORDS, ICD-9, MULTI-LABEL CLASSIFICATION, RECURRENT NEURAL NETWORKS, CONVOLUTIONAL NEURAL NETWORKS

In this master thesis, the problem of automatic tagging of electronic medical documents is being studied. The problem of code assignment (codes of diseases according to ICD-9) is solved according to discharge summaries.

The purpose of the research is to formalize the problem of automatic tagging of documents, to study and compare various approaches for its solution.

The results of the master thesis:

1. The corpus of documents was formed (about 53 000 cases of admissions)
2. Vector models of the corpus of documents are constructed, different approaches are compared
3. The models of multi-label classification have been trained and analyzed, the quality and complexity of models for datasets with different number of labels (50 and 1612)

The experiments are implemented in the Python programming language.

The master thesis was written solely by the author.