

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра биомедицинской информатики

Аннотация к дипломной работе

**«Алгоритм динамической нормализации и дополнения
биомедицинских заметок»**

Белясов Станислав Вячеславович

Научный руководитель – старший преподаватель,
Николаев Г. И.

Консультант – старший преподаватель,
Толстиков А. А.

Минск, 2022

РЕФЕРАТ

Дипломная работа, 56 стр., 32 иллюстр., 5 таблиц, 43 источника.

Ключевые слова: ВИЗУАЛИЗАЦИЯ, НАБОР ДАННЫХ, НЕЙРОННЫЕ СЕТИ, МЕХАНИЗМ ВНИМАНИЯ, РАНЖИРОВАНИЕ, ОБРАБОТКА ТЕКСТОВ, МАШИННОЕ ОБУЧЕНИЕ, ВЕКТОРНОЕ ПРЕДСТАВЛЕНИЕ ТЕКСТОВ, СТАТИСТИЧЕСКИЙ АНАЛИЗ.

Объект исследования — набор данных биомедицинских заметок, задача распознавания именованных объектов, задача нормализации именованных объектов.

Цель работы — распознавание и нормализация биомедицинских названий методами машинного обучения.

Методы исследования — компьютерная лингвистика, обработка естественного языка, методы предобработки текстов, методы токенизации текстов, методы получения векторных представлений, ранжирование.

Работа посвящена написанию алгоритма распознавания и нормализации биомедицинских названий в тексте и их визуализация в графическом интерфейсе.

РЭФЕРАТ

Дыпломная праца, 56 с., 32 рыс., 5 табліц, 43 крыніцы.

Ключавыя словы: ВІЗУАЛІЗАЦЫЯ, НАБОР ДАДЗЕНЫХ, НЕЙРО-НАВЫЯ СЕТКІ, МЕХАНІЗМ УВАГІ, РАНЖЫРАВАННЕ, АПРАЦОЎКА ТЭКСТАЎ, МАШЫННАЕ НАВУЧАННЕ, ВЭКТАРНАЯ ЎЯЎЛЕННЕ ТЭКСТАЎ, СТАТЫСТЫЧНЫ АНАЛІЗ.

Аб'ект даследавання — набор дадзеных біямедыцынскіх нататак, задача распазнавання найменных аб'ектаў, задача нармалізацыі найменных аб'ектаў.

Мэта работы — распазнаванне і нармалізацыя біямедыцынскіх назваў метадамі машыннага навучання.

Метады даследавання — кампутарная лінгвістыка, апрацоўка натуральнай мовы, метады предобработки тэкстаў, метады токенизации тэкстаў, метады атрымання вектарных уяўленняў, ранжыраванне.

Праца прысвечана напісанню алгарытму распазнавання і нармалізацыі біямедыцынскіх назваў у тэксце і іх візуалізацыя ў графічным інтэрфейсе.

ABSTRACT

Diploma thesis, 56 p., 32 fig., 5 tables, 43 sources.

Keywords: VISUALIZATION, DATA SET, NEURAL NETWORKS, ATTENTION MECHANISM, RANKING, TEXT PROCESSING, MACHINE LEARNING, VECTOR REPRESENTATION OF TEXTS, STATISTICAL ANALYSIS.

The object of research is a data set of biomedical notes, the task of recognizing named objects, the task of normalizing named objects.

Objective: recognition and normalization of biomedical names by machine learning methods.

Research methods — computational linguistics, natural language processing, text preprocessing methods, text tokenization methods, methods for obtaining vector representations, ranking.

The work is devoted to writing an algorithm for recognizing and normalizing biomedical names in the text and their visualization in the graphical interface.