

АННОТАЦИЯ

В научно-исследовательской работе рассматривается создание и возможность применения программного инструмента для определения и унификации названий научных организаций из научных публикаций. В работе проведен обзор существующих алгоритмов и технологий обработки аффилиаций, определены целевые составляющие аффилиаций, составлена схема базы данных определения и унификации наименований научных организаций, разработан и протестирован на точность алгоритм. На основе тестирования установлено, что с использованием программного инструмента возможно программное определение и унификация наименований научных организаций.

ABSTRACT

The research paper discusses the development and potential use of a software tool for identifying and standardizing the names of academic institutions from scholarly publications. The study provides an overview of current algorithms and techniques for processing affiliation data, identifies key components of affiliations, creates a database structure for determining and standardizing academic institution names, and develops and evaluates an accuracy algorithm. Based on testing, the paper concludes that using a software application, it is possible to automate the process of identifying and standardizing academic institutions' names.

СОДЕРЖАНИЕ

АННОТАЦИЯ	3
ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ	4
Введение	7
РАЗДЕЛ 1. МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ АФФИЛИАЦИЙ НАУЧНЫХ ОРГАНИЗАЦИЙ	9
1.1 Составные части аффилиаций	9
1.2 Анализ существующих решений	12
1.2.1 Google Geocoding API	12
1.2.2 Яндекс API «Поиск по организациям»	13
1.2.3 Описание OpenStreetMap	14
1.2.4 Итог анализа существующих решений	15
1.3 Анализ существующих нейросетевых решений	16
РАЗДЕЛ 2. РАЗРАБОТКА АЛГОРИТМА ИНСТРУМЕНТА ПО УНИФИКАЦИИ НАЗВАНИЙ НАУЧНЫХ ОРГАНИЗАЦИЙ	24
2.1 Алгоритм обработки и унификации названий научных организаций	24
2.2 Разработка алгоритма	27
2.3 Результаты работы алгоритма	35
2.3.1 Оценка качества работы алгоритма	40
ЗАКЛЮЧЕНИЕ	49
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	52