

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ

Кафедра теории вероятностей и математической статистики

КАЛГАНОВ Максим Андреевич

**ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ
РЕШЕНИЯ ЗАДАЧ NLU**

Аннотация к магистерской диссертации
специальность 1-31 80 09 «Компьютерный анализ данных»

Научный руководитель

Сталевская С.Н.

кандидат физико-математических наук,
доцент кафедры математического
моделирования и анализа данных

Минск, 2021

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Магистерская диссертация, 90 с., 22 рис., 22 источника

ПОНИМАНИЕ ЕСТЕСТВЕННОГО ЯЗЫКА, ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА, МАШИННОЕ ОБУЧЕНИЕ, МЕТОДЫ ГЛУБОКОГО ОБУЧЕНИЯ, ЛЮДВИГ ВИТГЕНШТЕЙН.

Объект исследования – методы машинного обучения и методы глубокого обучения, направленные на решение задач обработки естественного языка, задачи реализации механизма понимания.

Цель работы – проанализировать проблематику понимания естественного языка, и того, насколько исследователи в сфере обработки и понимания естественного языка смогли приблизиться к решению данной задачи. Предложить возможное понимание задачи и направления для её решения. Также реализовать одно из возможных решений задачи.

Результат – формализованная задача, обученная модель с проанализированным результатами обучения, описанные перспективы развития работы.

Магистерская диссертация состоит из четырех глав. В главе 1 производится обзор методов обработки естественного языка, рассматриваются основные подходы в решении задач NLP. Глава 2 посвящена проблематике реализации механизма понимания, анализируются методы, рассмотренные в первой главе, производится поиск ответов на концептуальные вопросы для точной формулировки задачи. В главе 3 формализуется задача, описывается ближайшая цель, которую необходимо достичь для реализации понимания, также вводится математическая модель для подзадачи. Глава 4 посвящена реализации описанной подзадачи, произведен анализ результатов, сделаны выводы.

АГУЛЬНАЯ ХАРАКТАРЫСТЫКА РАБОТЫ

Магістарская дысертацыя, 90 с., 22 мал., 22 крыніцы

РАЗУМЕННЕ МОВЫ, АПРАЦОЎКА МОВЫ, МЫШЫННАЕ НАВУЧАННЕ, МЕТАДЫ ГЛЫБОКАГА НАВУЧАННЯ, ЛЮДВІГ ВІТГЕНШТЭЙН.

Аб'ект даследавання - метады машыннага навучання і метады глыбокага навучання, накіраваныя на рашэнне задач апрацоўкі мовы, задачы рэалізацыі механізму разумення.

Мэта работы - прааналізаваць праблематыку разумення мовы, і таго, наколькі даследчыкі ў сферы апрацоўкі і разумення мовы змаглі наблізіцца да вырашэння дадзенай задачы. Прапанаваць магчымае разуменне задачы і напрамкі для яе рашэння. Таксама рэалізаваць адно з магчымых рашэнняў задачы.

Вынік - фармалізаваная задача, навучаная мадэль з прааналізаваным выніком навучання, апісаныя перспектывы развіцця работы.

Магістарская дысертацыя складаецца з чатырох частак. У главе 1 вырабляецца агляд метадаў апрацоўкі мовы, разглядаюцца асноўныя падыходы ў вырашэнні задач NLP. Частка 2 прысвечана праблематыцы рэалізацыі механізму разумення, аналізуюцца метады разгледжання ў першай частцы, вырабляецца пошук адказаў на канцэптואльныя пытанні для дакладнай фармулёўкі задачы. У раздзеле 3 фармалізуе задача, апісваецца бліжэйшая мэта, якую неабходна дасягнуць для рэалізацыі разумення, таксама ўводзіцца матэматычная мадэль для падзадачи. Частка 4 прысвечана рэалізацыі апісанай падзадачи, выраблены аналіз вынікаў, зроблены высновы.

GENERAL CHARACTERISTIC OF THE WORK

Master's thesis, 90 p., 22 figures, 22 sources

NATURAL LANGUAGE UNDERSTANDING, NATURAL LANGUAGE PROCESSING, MACHINE LEARNING, DEEP LEARNING, LUDWIG WITGENSTEIN.

Object of the research - machine learning methods and deep learning methods aimed at solving problems of natural language processing, tasks of implementing the understanding mechanism.

Research goal - to analyze the problems of understanding natural language, and results, which were reached by researchers in the field of processing and understanding natural language. Offer a possible understanding of the problem and directions for solving it. Also implement one of the possible solutions to the problem.

Result - formalized task, a trained model with analyzed learning outcomes, described prospects for the development of work.

The master's thesis consists of four chapters. Chapter 1 provides an overview of natural language processing techniques and the main approaches to solving NLP problems. Chapter 2 is devoted to the problems of the implementation of the understanding mechanism, the methods considered in the first chapter are analyzed, the search for answers to conceptual questions is made for the exact formulation of the problem. Chapter 3 formalizes the problem, describes the immediate goal that must be achieved to realize the understanding, and introduces a mathematical model for the subproblem. Chapter 4 is devoted to the implementation of the described subtask, the analysis of the results and conclusions are made.