

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Санкт-Петербургский национальный исследовательский  
Академический университет Российской академии наук»  
Центр высшего образования

Кафедра математических и информационных технологий

Подгузов Никита Владимирович

# Определение правильного ответа на StackOverflow

Бакалаврская работа

Допущена к защите.  
Зав. кафедрой:  
д. ф.-м. н., профессор Омельченко А. В.

Научный руководитель:  
Курбанов Р. Э.

Рецензент:

-

Санкт-Петербург  
2018

SAINT-PETERSBURG ACADEMIC UNIVERSITY  
Higher education centre

Department of Mathematics and Information Technology

Nikita Podguzov

# Correct answer detection on StackOverflow

Graduation Thesis

Admitted for defence.  
Head of the chair:  
professor Alexander Omelchenko

Scientific supervisor:  
Rauf Kurbanov

Reviewer:  
-

Saint-Petersburg  
2018

# Оглавление

Введение	4
1. Обзор решений	6
2. Цели и задачи	7
3. Описание данных	8
4. Методы и реализация	9
4.1. Методы . . . . .	9
4.2. Реализация модели . . . . .	9
5. Обсуждение результатов	10
6. Заключение	11
Список литературы	12

# Введение

С распространением и развитием интернета стали популярны системы вопросов и ответов — вид сайтов, позволяющий пользователям задавать вопросы и отвечать на вопросы, заданные другими пользователями. Среди наиболее популярных подобных систем можно выделить многопрофильные (такие, как *Yahoo Answers*, *Quora*) и узкоспециализированные (например, *Stackoverflow*). В рамках данной работы мы сконцентрируемся именно на системе Stackoverflow, которая в первую очередь является системой для вопросов по программированию и смежным областям.

На Stackoverflow у пользователя есть возможность после того, как он задал интересующий его вопрос, отметить ответ, решивший его проблему, как правильный. Несмотря на это, достаточно большая часть вопросов остается без подобной отметки, что может усложнить поиск нужной информации по данному вопросу другим пользователям.

\*Сюда можно добавить скриншот вопроса с StackOverflow с интересующими нас частями веб-страницы\*

Таким образом появляется задача определения правильного ответа для вопросов с Stackoverflow.

В связи с многократно возросшими вычислительными мощностями в последнее время в области машинного обучения приобрели популярность нейронные сети. Сфера применимости моделей, использующих нейронные сети, очень широка: от распознавания образов на изображениях до предсказания курсов валют на бирже. Несмотря на наличие работ, исследующих задачу предсказания лучшего ответа, в основном в них используются более классические методы машинного обучения, хотя нейронные сети и показывают достойные результаты в задачах классификации текстов.

В данной работе исследуется применимость модели, основанной на нейронных сетях, к поставленной задаче, а также сравниваются ее модификации, использующие специфику задачи: технического языка вопросов, а также наличие сниппетов кода в вопросах и ответах.

## Кратко о последующих главах

В главе 1 производится обзор имеющихся решений, их преимуществ и недостатков.

В главе 2 описывается формальная постановка задачи, а также рассматриваются ее особенности и ставятся основные цели и задачи работы.

В третьей главе описана работа по предобработке и очистке данных с сайта StackOverflow, а также их анализ.

Глава 4 содержит описание основных подходов к поставленной задаче, а также детали их реализаций.

В последней главе проведен анализ полученных результатов и сделаны выводы о применимости модели к задаче определения правильного ответа.

# 1. Обзор решений

Существующие решения

## **2. Цели и задачи**

Цели

### **3. Описание данных**

Данные



## **4. Методы и реализация**

### **4.1. Методы**

Методы

### **4.2. Реализация модели**

Реализация

## **5. Обсуждение результатов**

Эксперименты

## **6. Заключение**

Заключение

## **Список литературы**