

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к магистерской диссертации

**ВОССТАНОВЛЕНИЕ ПРОПУСКОВ В ТЕКСТЕ С ПОМОЩЬЮ
СТАТИСТИЧЕСКИХ ЯЗЫКОВЫХ МОДЕЛЕЙ**

Сологуб Анастасия Сергеевна

Научный руководитель — кандидат физико-математических наук,
доцент И. А. Бодягин

2015

Реферат

Магистерская диссертация, 32 страницы, 5 рисунков, 19 источников, 1 приложение.

ЯЗЫКОВОЕ МОДЕЛИРОВАНИЕ, N-ГРАММЫ, НЕЙРОННАЯ СЕТЬ ПРЯМОГО РАСПРОСТРАНЕНИЯ, РЕКУРРЕНТНАЯ НЕЙРОННАЯ СЕТЬ, МОДЕЛЬ МАКСИМАЛЬНОЙ ЭНТРОПИИ, МЕТОД РАСПРОСТРАНЕНИЯ ОШИБКИ СКВОЗЬ ВРЕМЯ.

Объект исследования — языковые модели.

Цель работы — построить языковую модель для восстановления пропущенного слова в предложении.

Методы исследования – методы языкового моделирования.

Результат – программная модель для восстановления пропущенного слова.

Область применения – задачи обработки естественного языка.

Структура диссертации: диссертация состоит из введения, двух глав и заключения. В первой главе приводится обзор основных статистических моделей, использующихся для языкового моделирования и методов их применения к задаче восстановления пропусков в тексте. Во второй главе представлены результаты практического эксперимента применения модели рекуррентной нейронной сети для восстановления пропущенного слова.

Abstract

Master thesis, 32 pages, 5 figures, 19 reference, and 1 appendix.

LANGUAGE MODELING, N-GRAMS, FEEDFORWARD NEURAL NET, RECURRENT NEURAL NET, MAXIMUM ENTROPY MODEL, BACKPROPAGATION THROUGH TIME.

The object of the study – language models.

The purpose of work – development of language model for imputing missing word at the sentence.

Research methods – language modeling methods.

The result is the language model for inserting back the missing word.

The application field is natural language processing applications.

Thesis consists of introduction, two chapters and conclusion. First chapter contains description of cutting-edge language models and their application to imputing missing words. Second chapter contains results of using recurrent neural network for missing position detecting and word insertion.