

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ РАДИОФИЗИКИ И КОМПЬЮТЕРНЫХ
ТЕХНОЛОГИЙ

Кафедра телекоммуникаций и информационных технологий

МОЖЕЙКО
Владислав Павлович

РАЗРАБОТКА АЛГОРИТМА ПОСТРОЕНИЯ ВИЗУАЛЬНЫХ
ФОРМ
НА ОСНОВЕ СЕМАНТИЧЕСКОГО АНАЛИЗА ТЕКСТА

Аннотация к дипломной работе

Научный руководитель – кандидат техн. наук,
доцент К.С. Мулярчик

Минск, 2022

РЕФЕРАТ

Дипломная работа: 61 с., 17 рис., 1 табл., 18 источников, 4 прил.

ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА, РАСПОЗНАВАНИЕ ИМЕНОВАННЫХ СУЩНОСТЕЙ, ПОСТРОЕНИЕ ВИЗУАЛЬНЫХ ФОРМ, КЛАССИФИКАЦИЯ СООБЩЕНИЙ

Цель работы – разработка алгоритмов автоматизированного построения визуальных форм для сообщений электронной почты с использованием семантического анализа текста.

Объект исследования – базовые составляющие визуальных форм и алгоритмы построения визуальных форм.

Метод исследования: классификация базовых элементов визуальной формы, использование нейронных сетей классификации текста и распознавания именованных сущностей, анализ связей между словами на базе синтаксических деревьев.

В ходе работы реализованы алгоритмы построения визуальных форм на базе разработанной классификации. Использована библиотека Spacy для анализа связей между словами в тексте. Для классификации токенов и сообщений использовалась дообученная нейронная сеть трансформатор BERT. В помощью векторной графики, рисовались визуальные формы на основе полученной в ходе обработки информации

Разработанная система может быть использована как вспомогательная программа при анализе большого количества email сообщений.

РЭФЕРАТ

Дыпломная работа: 61 с., 17 мал., 1 таб., 18 крыніц, 4 дад.

АПРАЦОЎКА НАТУРАЛЬнай мовы, РАЗАВЫЧАННЕ ІМЯВАНых СУТНАСЦЯЎ, ПАБУДАВАННЕ ВІЗУАЛЬНых ФОРМ, КЛАСІФІКАЦЫЯ ПАВЕДАМЛЕННЯЎ

Мэта работы – распрацоўка алгарытмаў аўтаматызаванага пабудовы візуальных форм для паведамленняў электроннай пошты з выкарыстаннем семантычнага аналізу тэксту.

Аб'ект даследавання – базавыя складнікі візуальных форм і алгарытмы пабудовы візуальных форм.

Метад даследавання: класіфікацыя базавых элементаў візуальнай формы, выкарыстанне нейронавых сетак класіфікацыі тэксту і распазнанні найменных сутнасцяў, аналіз сувязяў паміж словамі на базе сінтаксічных дрэў.

У ходзе работы рэалізаваны алгарытмы пабудовы візуальных форм на базе распрацаванай класіфікацыі. Выкарыстана бібліятэка Spacy для аналізу сувязей паміж словамі ў тэксце. Для класіфікацыі токенаў і паведамленняў выкарыстоўвалася данавучаная нейронавая сетка трансфарматар BERT. З дапамогай вектарнай графікі маляваліся візуальныя формы на аснове атрыманай падчас апрацоўкі інфармацыі

Распрацаваная сістэма можа быць скарыстана як дапаможная праграма пры аналізе вялікай колькасці email паведамленняў.

ABSTRACT

Thesis: 61 p., Figures 17, Table 1, 18 sources, 4 app.

NATURAL LANGUAGE PROCESSING, NAMED ENTITY
RECOGNITION, VISUAL FORMS CONSTRUCTION, MESSAGE
CLASSIFICATION

Objective – development of algorithms for automated construction of visual forms for e-mail messages using semantic text analysis.

Research object – basic components of visual forms and algorithms for visual form constructing.

Research methods: classification of the basic elements of a visual form, using of neural networks for text classification and named entity recognition, analysis of relationships between words based on syntax dependency trees.

In the course of the work, algorithms for constructing visual forms based on the developed classification were implemented. The Spacy library was used to analyze the relationships between words in the text. To classify tokens and messages, a fine-tuned neural network BERT transformer was used. Using vector graphics, visual forms were drawn based on the information received during processing.

The developed system can be used as an auxiliary program when analyzing a large number of email messages.