

КУЗНЕЦОВА АНАСТАСИЯ

+7 926 084 3731 ♦ ferreira.a-ferreira@yandex.ru

ОБРАЗОВАНИЕ

НИУ Высшая Школа Экономики
Магистратура
Компьютерная лингвистика

Сентябрь 2017 - Наст. время
GPA: 8.98 (10-балльная шкала)

Российский Государственный Гуманитарный Университет
Бакалавриат

Сентябрь 2013 - Июль 2017
GPA: 4.98 (5-балльная шкала)

КАРЬЕРНЫЕ ЦЕЛИ

Получить опыт решения практических NLP задач, решаемых в production, на реальных данных. Профессионально развиваться и быть в курсе последних тенденций и исследований в области обработки текстов на естественном языке.

ТЕКУЩИЕ ПРОЕКТЫ

Named Entity Recognition для корпуса русских научно-популярных текстов

Цель исследовательского проекта – изучить научно-популярные тексты на русском языке с помощью методов компьютерной лингвистики. Моей задачей было извлечение из текстов имен ученых и создание веб-интерфейса с помощью Django. Имена извлекались с использованием технологии Яндекса Томита-парсер. Достигнута ассигатура NER-парсера на указанном корпусе 0.48.

Разрешение морфологической неоднозначности в парагваском гуарани

Разрешение неоднозначности в гуарани (язык в Южной Америке) осуществляется при помощи системы правил и статистически выровненного параллельного корпуса с использованием НММ и IBM алгоритмов.

Система машинного перевода для пары гуарани–испанский

Проект стажировки Google Summer of Code для open-source платформы машинного перевода Apertium, специализирующейся на малоресурсных языках. Наиболее значимым результатом проекта стало покрытие (coverage) морфологическим анализатором 0.89 и 0.91 токенов на двух разных корпусах.

Link Prediction в графе параллельного корпуса гуарани-испанский.

Задача состояла в классификации ребер графа параллельного выровненного по предложениям корпуса гуарани-испанский. F-мера составила 0.97 при использовании логистической регрессии.

ТЕХНИЧЕСКИЕ НАВЫКИ

Языки программирования
Библиотеки & Фреймворки
Основные навыки

Python, HTML, Bash, R
Skearn, NLTK, Pandas, Keras, HFST, Django
Fact Extraction, Sentiment analysis, KeywordExtraction,
Vectorization

ОПЫТ РАБОТЫ

Конкурс Google Code-In
Ментор

2018

- Менторство для платформы машинного перевода Apertium. Помощь участникам конкурса в задачах, связанных с работой в UNIX-системах и программировании.

Стажировка Google Summer of Code

2018

АКАДЕМИЧЕСКИЕ УСПЕХИ

Повышенная Государственная Академическая Стипендия (за достижения в научно-исследовательской деятельности) 2018 (НИУ ВШЭ).

Повышенная Государственная Академическая Стипендия (за достижения в научно-исследовательской деятельности): 2 семестр 2014/2015 учебного года, 1 и 2 семестры 2015/2016 учебного года, 1 семестр 2016/2017 учебного года (РГГУ)

ДРУГОЕ

2014 – 2017 Волонтерская работа в Бразильском культурном центре в Москве. Чтение публичных лекций по истории и культуре Бразилии, преподавание португальского языка.

Изучение иностранных языков: Английский (TOEFL 103, Advanced), Португальский (CELPE-BRAS, Upper Intermediate), Испанский (Upper Intermediate), Литовский (Upper Intermediate).

ДОПОЛНИТЕЛЬНЫЕ КУРСЫ

Introduction to Probability and Data, Duke University (Grade Achieved: 90.5%)¹

Inferential Statistics, Duke University (Grade Achieved: 93.0%)²

¹<https://www.coursera.org/account/accomplishments/certificate/FUCGZH2HS836>

²<https://www.coursera.org/account/accomplishments/certificate/5TDEB6YTS7U>