



Пичужкина Ольга Валерьевна

Интересы

Python, UNIX, data engineering, data science, ML, NLP, deep learning, digital humanities

Ключевые навыки

Python, C, UNIX, SQL, HTML, CSS, data science, machine learning, deep learning, NLP

Языки

English

(*advanced*, 8 за внутренний экзамен НИУ ВШЭ по английскому языку, свободно говорю по-английски, есть опыт написания и редактирования профессиональных текстов)

Deutsch

(*intermediate [B2]*)

Italian, Español, Portuguese

(*elementary [A2]*)

Образование

09.2016 – ОП бакалавриата «Фундаментальная и компьютерная лингвистика», НИУ ВШЭ, Москва.

06.2021 Разработка на Python, ООП, алгоритмы и структуры данных, natural language processing.

01.2017 – Майнор «Интеллектуальный анализ данных» факультета компьютерных наук, НИУ ВШЭ, Москва.

Data science на Python, machine learning, deep learning, natural language processing.

03.2019 – Школа программирования 21, ПАО Сбербанк, Москва, уровень 6.80.

07.2020 Разработка на C, UNIX, алгоритмы и структуры данных, принципы написания чистого и эффективного кода.

08.2018 Курс Neural Networks and Deep Learning, [deeplearning.ai](#), (сертификат).

01.2019 Курс Introduction to Deep Learning, НИУ ВШЭ, (сертификат).

09.2020 Курс Основы HTML и CSS, Нетология-групп, (сертификат).

Инструменты/фреймворки

ЯП Python, C, go lang, R, shell scripts (bash, zsh)

Анализ данных Excel, SQL (MySQL, SQLite, PostgreSQL), json, pandas, numpy, matplotlib, seaborn, plotly, scikit-learn, scipy, tensorflow, keras, keras_bert, pytorch, математическая статистика

Типография и веб HTML, CSS, LaTeX, Markdown, Django, Flask, urllib, requests, BeautifulSoup

NLP NLTK, gensim, word embeddings (Word2Vec, FastText, ELMO, BERT), регулярные выражения, pymorphy2, mystem (pymystem3), NER, CoNLL-U

Дополнительно

01.2021

Участвовала в хакатоне «Fear and Loathing in DH» Центра DH-исследований НИУ ВШЭ. Результат.

10.2020 — 12.2020

Есть опыт ведения telegram-канала через бота-администратора, привязанного к каналу с помощью Controller Bot.

04.2020 — 08.2020

Бета-тестировала задачи интенсива по Python и ML «Школы 21» («Piscine Python for Data Science»).

10.2019

Участвовала в хакатоне программы стажировок SafeBoard «Лаборатории Касперского». Результат – была приглашена на обучение по направлению «Разработка».

Проекты

Python

- **«Прожито»: topic modelling**
Тематическое моделирование корпуса текстов: связь тем дневниковых записей с историческими событиями времени написания.
- **Quora question search**
Приложение на django – поисковик на нескольких моделях поиска, осуществляющий поиск по коллекции документов.
- **20 Newsgroups ciphertext classification**
Классификатор, обученный на датасете, состоящем из зашифрованных новостных текстов. Финальный макро f1-score в соревновании на kaggle — 0.49.
- **IMDB review classification**
Нейросеть, обученная на корпусе отзывов с IMDB и предсказывающая по тексту отзыва поставленный фильму рейтинг.
- **deepdream.ipynb (Google Colab)**
Эксперименты с технологией DeepDream, выполненные на фреймворке tensorflow.
- **Курсовая работа «Автоматическое обнаружение слов русского языка с короткой историей (на материале газетных текстов последнего десятилетия)»**
- **Gotham_bot (telegram-бот)**
Марковская цепь, генерирующая текст на основе сценария сериала «Готэм».
- **Automatic extraction and correction of bad collocations in academic writing**
Скрипт для поиска лексически/стилистически неправильных сочетаний (коллокаций) в академических текстах и предложения вариантов замены.

C

- **minishell**
Мини-интерпретатор командной строки UNIX.
- **ft_select**
Графический интерфейс для терминала, позволяющий передвигаться по нему стрелочками и выделять объекты.
- **FdF**
Программа, которая рендерит 3-D поверхность по двумерной карте.
- **ft_ls**
Своя реализация утилиты UNIX *ls*.
- **push_swap**
Алгоритм, сортирующий заданную последовательность минимальным количеством заранее определённых операций.
- **ft_printf**
Своя реализация функции, парсящей форматированную печать.
- **fillit**
Алгоритм, составляющий из заданных фигурок тетриса квадрат наименьшего размера.
- **libft + get_next_line**
Своя реализация некоторых функций из стандартной библиотеки C и несколько дополнительных функций.

Опыт работы

- 01.21 – н. вр. **ML-разработчик, IT For You, Москва.**
Решение ML-задач: классификация звуков и шумов органического происхождения.
- 02.19 – 08.19, **Автор, hyperskill.org (Jetbrains Academy).**
01.20 – н. вр. *Составление на английском языке текстов, задач и тестов к задачам на обучающий сайт в разделы «Math», «Python» и «Essentials» (полный список разработанных материалов).*
- 09.19 – н. вр. **Разметчик данных, АВВУУ, Москва.**
Разметка именованных сущностей в документах на русском, английском, немецком, испанском, итальянском и португальском языках.
- 05.20 – н. вр. **Модератор комментариев, hyperskill.org (Jetbrains Academy).**
Разбор комментариев в разделах «Algorithms» и «Math» обучающего сайта, исправление ошибок/опечаток в текстах и задачах, на которые жалуются комментаторы, разработка материалов курса.
- 04.20 – 07.20 **Разметчик данных, ПАО Сбербанк, Москва.**
- 09.19 – 12.19, **Сопровождающий онлайн-курса Введение в искусственный интеллект, НИУ ВШЭ, Москва.**
03.20 – 08.20, *Ответы на вопросы студентов, решение технических проблем, разработка материалов курса.*
- 09.20 – н. вр.
- 09.19 – 07.20 **Учебный ассистент по предмету «Программирование и лингвистические данные» у студентов 1 курса ОП «Фундаментальная и компьютерная лингвистика» (сайт курса), НИУ ВШЭ, Москва.**
Ответы на вопросы студентов, проверка работ, разработка материалов курса.
- 06.18 – 02.20 **Разметчик данных, Yandex, Москва.**
- 04.19 - 06.19 **Учебный ассистент по предмету «Введение в анализ данных» у студентов 1 курса бакалавриата МирЭк, НИУ ВШЭ, Москва.**
Ответы на вопросы студентов, проверка работ.
- 08.18 – 12.18, **Учебный ассистент по предмету «Компьютерные инструменты лингвистических исследований» у студентов 1 курса ОП «Фундаментальная и компьютерная лингвистика», НИУ ВШЭ, Москва.**
09.17 – 12.17 *Ответы на вопросы студентов, проверка работ, разработка материалов курса.*