

Social issue is just a

World

AI & Data

Challenge

МЕЖДУНАРОДНЫЙ КОНКУРС

ПО РЕШЕНИЮ ГЛОБАЛЬНЫХ

СОЦИАЛЬНЫХ ЗАДАЧ И ВНЕДРЕНИЮ ИХ

РЕШЕНИЙ ПРИ ПОДДЕРЖКЕ

МЕЖДУНАРОДНЫХ ЭКСПЕРТОВ



Automating Total Dictation

Индивидуальный участник

Кот Николай Валериевич, кандидат физико-математических наук

<https://git.asi.ru/Kot/2-automating-total-dictation-verification>

2. Создание алгоритма автоматической проверки «Тотального диктанта»

Все субъекты РФ – Тотальный диктант

Задачи проекта:

1. Интегрируемый модуль или скрипт, который можно встраивать на наш сайт в разные разделы: собственно раздел с онлайн-диктантом, страницы с уроками онлайн-школы, в коллекцию диктантов. Продукт нужно будет интегрировать также не только на сайт диктанта, но в разные образовательные продукты (на сайтах, образовательных платформах, в мобильных приложениях).
2. После написания диктанта должна быть возможность проанализировать тепловую карту ошибок и написания, внести корректировки в алгоритм проверки, добавить неучтенные моменты.
3. Скрипт должен показывать ошибки в тексте, при наведении во всплывающем окне показывать правильное написание, варианты написания и комментарии должна быть возможность ранжировать по приоритету.
4. Скрипт должен подсчитывать отдельно количество орфографических ошибок и пунктуационных, чтобы алгоритм на сайте мог выставить оценку. Описки и другие виды ошибок должны выделяться в тексте, но не отражаться на общем подсчете ошибок.
5. Ошибки в названии (в заголовке) не должны отражаться на общем подсчете ошибок.
6. Должна быть опция «пересчитывать» результат. То есть после внесения корректировок еще раз прогонять все написанные тексты через скрипт, чтобы выдать новый результат уже с учетом правок и дополнений.
7. Должна быть разработана среда (интегрируемая) для работы филологов, в которой они могут анализировать тепловую карту ошибок, вносить правки в алгоритм проверки, ранжировать варианты написания по приоритету.
8. На стороне филолога: должен быть создан интерфейс для разметки текста, накопление базы комментариев и возможность использовать уже готовые комментарии, пунктограммы, орфограммы для разметки будущих текстов.
9. Интерфейс должен быть адаптирован и для веб, и для мобильной версии.

Существующее уже решение с разметкой данных

После проверки диктанта пользователь получит отчёт с подробным объяснением каждой допущенной им ош/и дидактические материалы.

Старинный сад

Старинный сад

Скамейка была хорша: на двух волунах покоилась дубовое бривно, плавно и очень удобно углубленное в середине ; сзади и сверху скамьк деревянный столб. По крупным, с пр округлую каменную чашу , и пере разбегались к прибрежным заросл волных красно-желто-зеленые кор В этом сдау все было необычно: д стрижные кусты, прямо-угольный не видел никогда.

Ваши ошибки

Орфография 23

Пунктуация 4

Орфографическая ошибка

бривно

Написание буквы **е** на месте безударного гласного в корне проверяется подбором другой формы этого слова или однокоренного слова, в котором ударение падает на корень, например: *брёвна*. При проверке буквы **е** и **ё** взаимно заменяют друг друга.

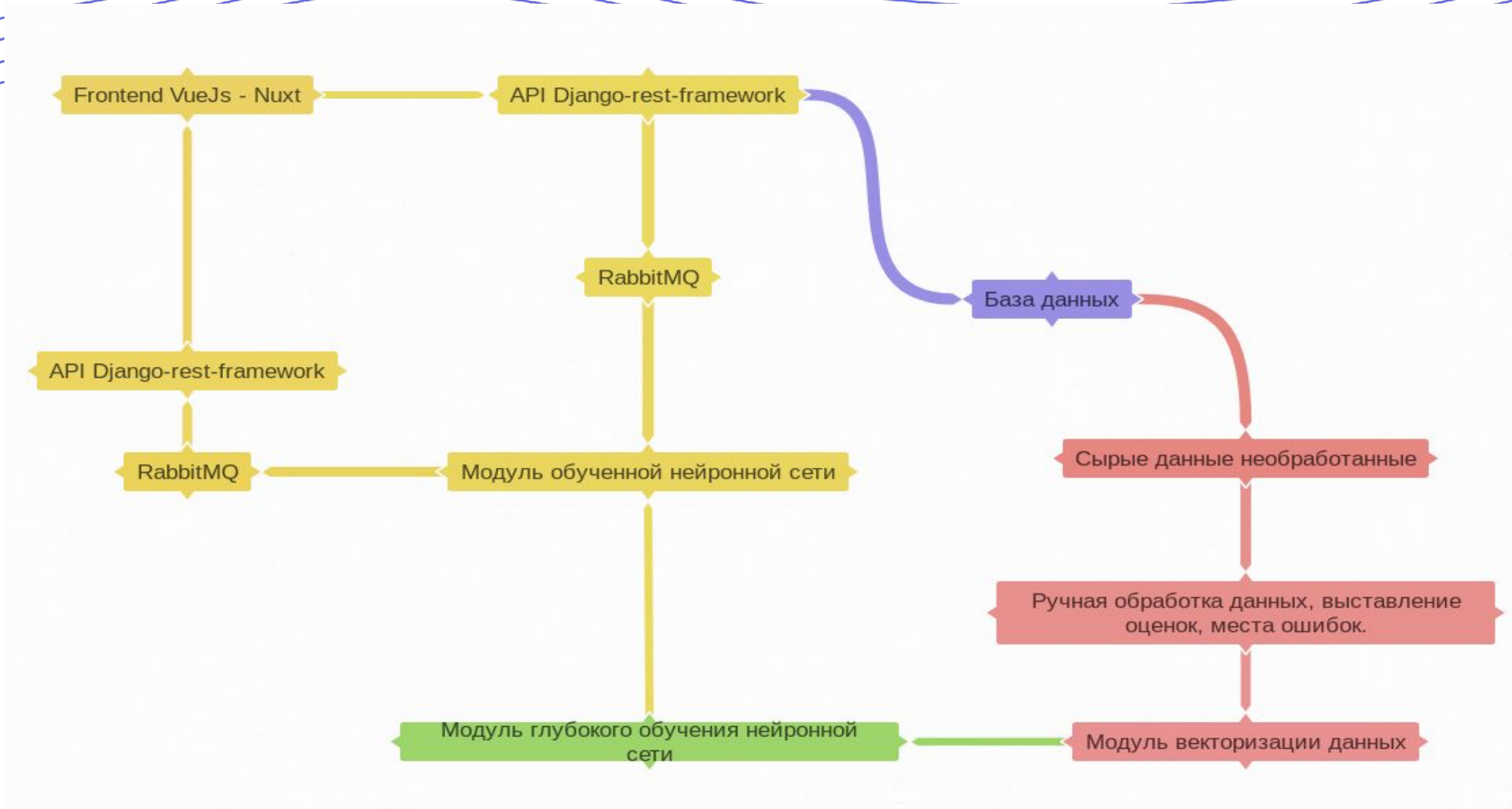
ПАС. Орфография. §34

Полногласные и неполногласные сочетания.
Безударные проверяемые гласные в корнях слов.
Безударные непроверяемые гласные.

орфография + 1

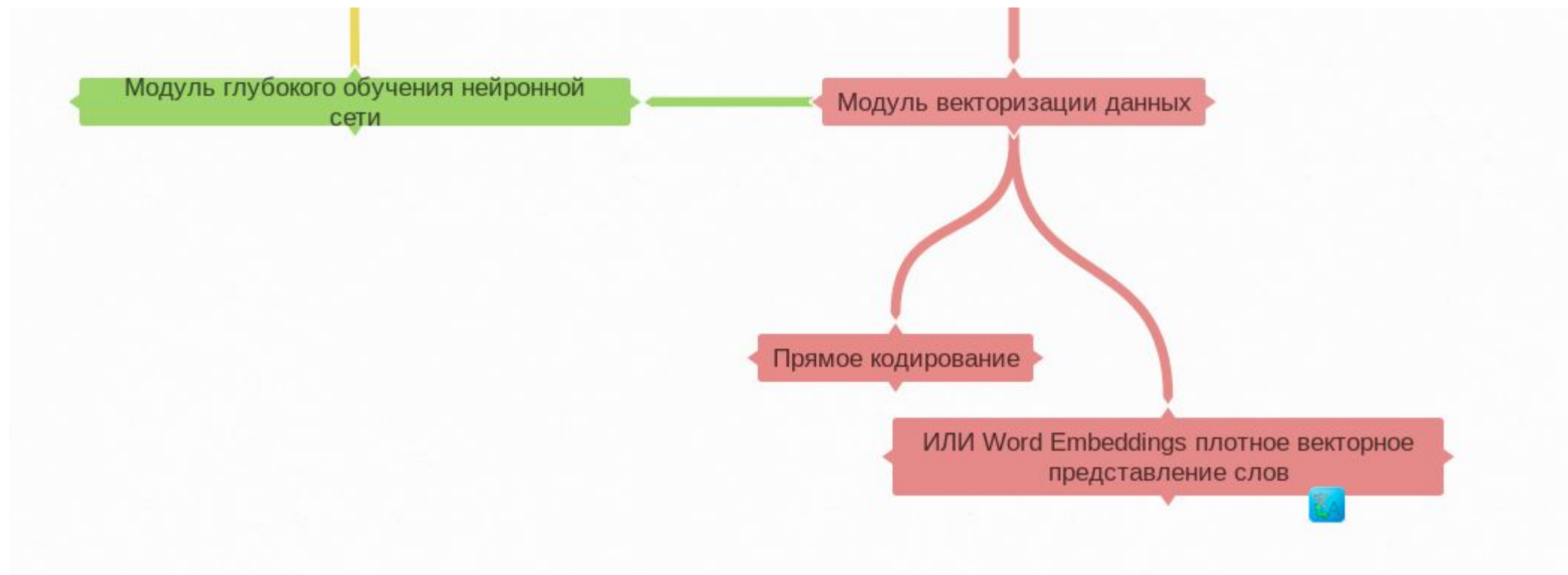
Описание решения.

1. Создание бекенд на Django-rest-framework для получения и анализа данных онлайн диктанта .
2. Отдельная часть для анализа поступивших данных на базе глубокого обучения нейронных сетей и возврата ответа.
3. Модуль для глубокого обучения нейронной сети на базе CNN сверточной нейронной сети либо рекуррентной нейронной сети RNN.
4. Модуль сбора и обработки данных для последующего обучения нейронной сети



<https://coggle.it/diagram/XvSe9k3nAhfkYCHG/t/-/4b3b159e850f4e5c3a447773166b4d7afab4cd58292f2071daa8cf6a12685526>

Возможно использование прямого кодирования слов либо Word Embeddings плотное векторное представление слов



Источником данных для обучения могут стать книги в электронном текстовом формате.



Есть еще много разных вариантов!

Есть множество шрифтов имитирующих рукописный текст. Это может быть использовано для обучения, если потребуется распознавание рукописного текста.

Fonts Online

Plaster of Paris

♡ 📄 43 K6

Fonts Online

Honey ScriptCyr Light

♡ 📄 96 K6

Fonts Online

Honey ScriptCyr Light

DreamsLiem

♡ 📄 156 K6

Fonts Online

yellosun

♡ 📄 40 K6

Fonts Online

mr_KindlyJasmineG

♡ 📄 77 K6

Fonts Online

mr_DopestyleG

♡ 📄 79 K6

Результат работы сети в виде тепловой карты активации класса, плюс оценка на выходе (скалярная или векторная регрессия)

После проверки диктанта пользователь получит отчёт с подробным объяснением каждой допущенной им ошибки и дидактические материалы.

Аналог этой штуки с фотографией слонов.

Старинный сад

Старинный сад
Скамейка была хорша: на двух волунах покоилась дубовое бривно, плавно и очень удобно углубленное в середине ; сзади и сверху скамьк...
деревянный столб. По крупным, с пр...
округлую каменную чашу , и пере...
разбегались к прибрежным заросл...
волных красно-желто-зеленые кора...
В этом сдау все было необычно: д...
стрижннные кусты, прямо-угольный...
не видел никогда.

Орфографическая ошибка

бривно

Написание буквы **е** на месте безударного гласного в корне проверяется подбором другой формы этого слова или однокоренного слова, в котором ударение падает на корень, например: **брёвна**. При проверке буквы **е** и **ё** взаимно заменяют друг друга.

ПАС. Орфография. §34

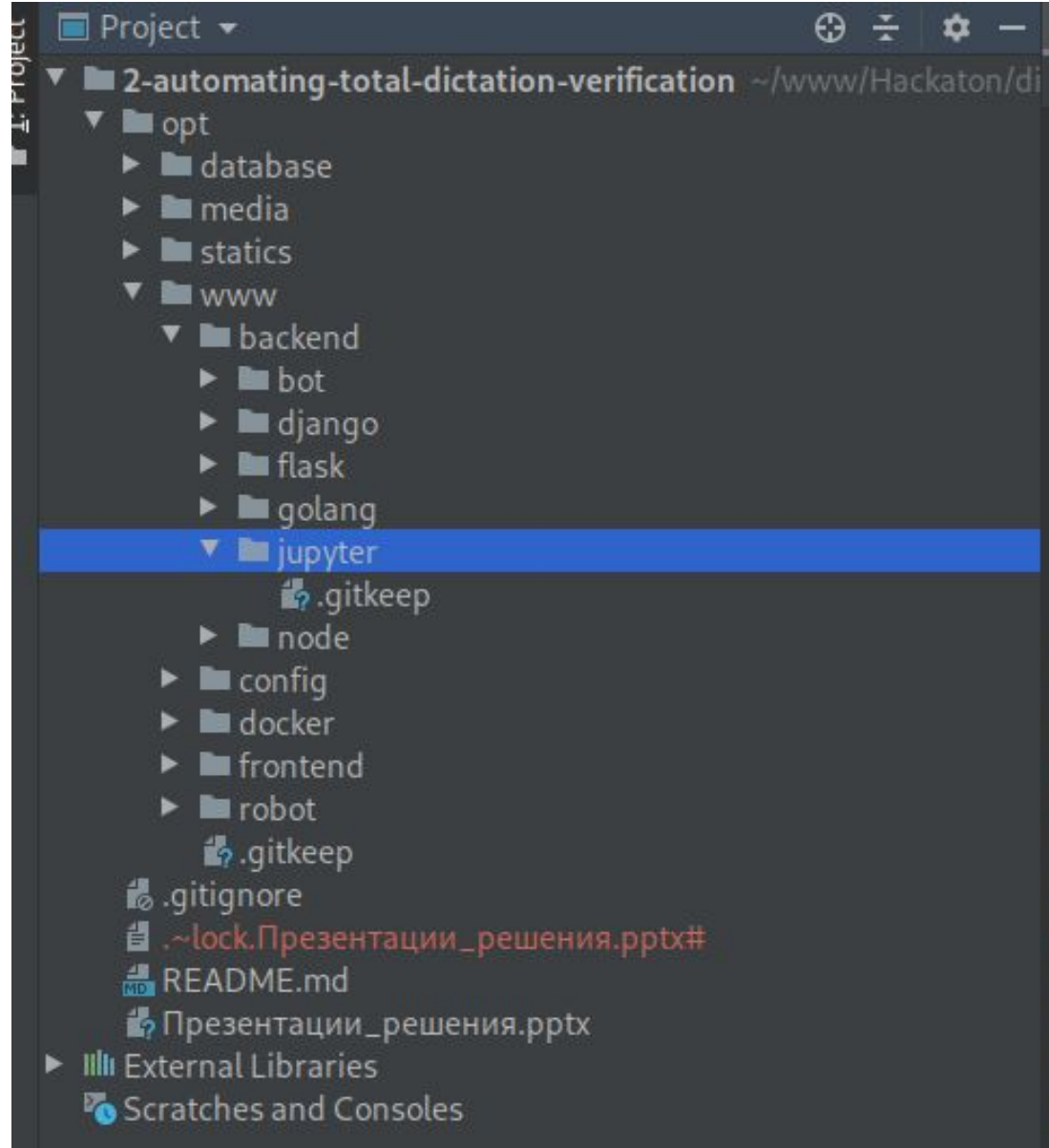
Полногласные и неполногласные сочетания.
Безударные проверяемые гласные в корнях слов.
Безударные непроверяемые гласные.

орфография + 1

Ваши ошибки	
Орфография	23
Пунктуация	4



Структура проекта





Использованные технологии:

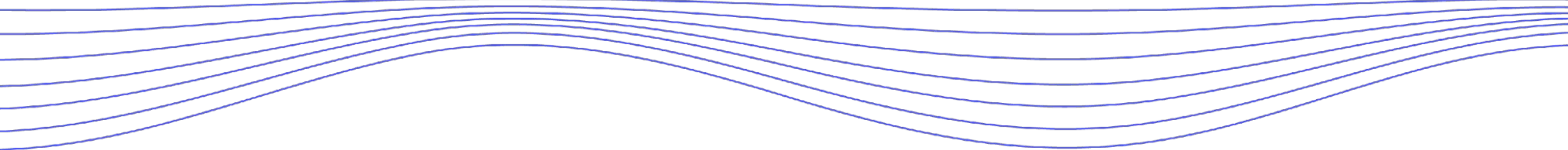
- Python, Django-rest-framework, Vuejs+Nuxt
- Keras, tensorflow, или возможно pytorch
- Linux, Docker, docker-compose



Итоги решения.

Результативность: программное решение позволит достигнуть поставленных в задаче целей (слайд3), а также может быть использовано в онлайн обучении для онлайн тестов, онлайн диктантов в онлайн школах что сейчас является особо актуальной темой.

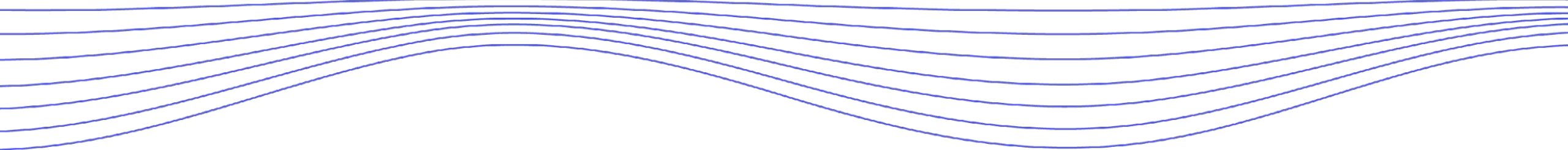
Выводы: данная тема является малоизученной но актуальной. Решение позволит автоматизировать процессы онлайн диктанта и выходящего на передний план онлайн образования.



Программное решение может быть использовано в действующих системах и решения реальных задач субъектов РФ, проект использует компоненты с открытым исходным кодом. Результат разработки, его исходный код возможно тоже может быть открыт.

Программное решение может быть масштабировано:

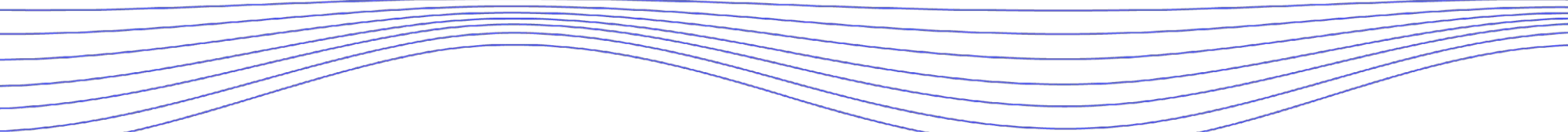
1. За счет использования docker контейнеров.
2. За счет возможности сохранения обученных моделей и последующего их повторного использования.



Для решений с UI: Решение имеет визуальную web часть, она основана на разработанном автором фреймворке wildleek. Пример работы можно посмотреть на видео одного из проектов автора:

<https://youtu.be/orCuTDmPV0I>

<https://youtu.be/wN0h8U6cAWI>



Готовность команды участвовать в реализации пилотных внедрениях своего решения в субъектах РФ в период с сентября 2020 по январь 2021.

Состав Команды

(Кот Николай Валериевич, Кандидат физико-математических наук, программист, фрилансер, самозанятый. <https://app.beeqb.com> <https://tranzilla.ru>
<https://nvkot.ru> <https://vk.com/nvkotru>
<https://vk.com/openschool26>)

Участник
Кот Николай Валериевич





D | A | T | A
I | A | T | A
M A S T E R S

datamasters.ru