



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Факультет экономики, менеджмента и
бизнес-информатики

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ ПОИСКА ИСПОЛНИТЕЛЕЙ ПО ТЕХНИЧЕСКОМУ ЗАДАНИЮ ПРИКЛАДНОГО ПРОЕКТА

Автор Соломатин Р. И.
Научный руководитель Бузмаков А. В.

Пермь, 2021



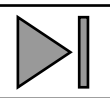
План доклада

1. Актуальность
2. Цели и задачи
3. Проектирование
4. Тесты
5. Результаты
6. Заключение



Актуальность

Каждый день выкладывается много тендеров. В ВШЭ много людей с разными компетенциями, и надо по текстовому описанию тендера понять, кто его сможет сделать.

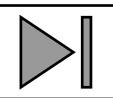




Проблема и гипотеза

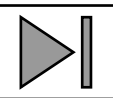
Проблема - автоматический поиск исполнителей под тендер, заданный текстовым описанием.

Гипотеза - можно ли автоматически подобрать исполнителя на основании анализа текстов ВКР.



Существующие решения

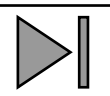
- Много тендеров
- Много людей с разными компетенциями
- Тяжело масштабировать, потому что необходимо знать много про разных людей
- Тратится много времени





Цель

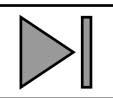
Цель работы – создать информационную систему для поиска исполнителя по текстовому описанию тендера.



Объект и предмет

Объект исследования - процесс поиска исполнителей по текстовому описанию тендера.

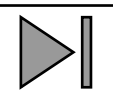
Предмет исследования - автоматизация процесса поиска исполнителей по текстовому описанию тендера.





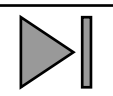
Задачи

- Сбор данных
- Разработать информационную систему



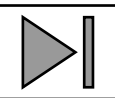


Проектирование



Выбор технологий

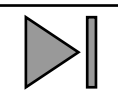
Язык программирования	Библиотека для МО	Предобученные модели	Работа с сайтом	Работа с файлами	Интерактивный режим
Python	++	++	+	+	+
C#	+	+	+	+	-
Java	+	+	+	+	-
JavaScript	+	+	+	+	?





База данных

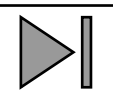
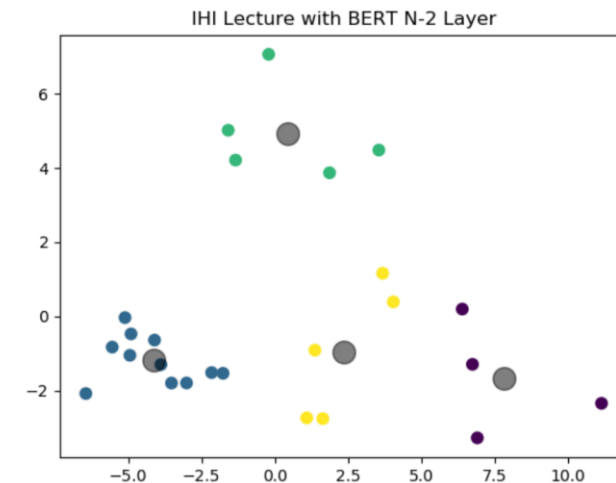
Данные брались со сайта
ВШЭ и помещались в базу
данных



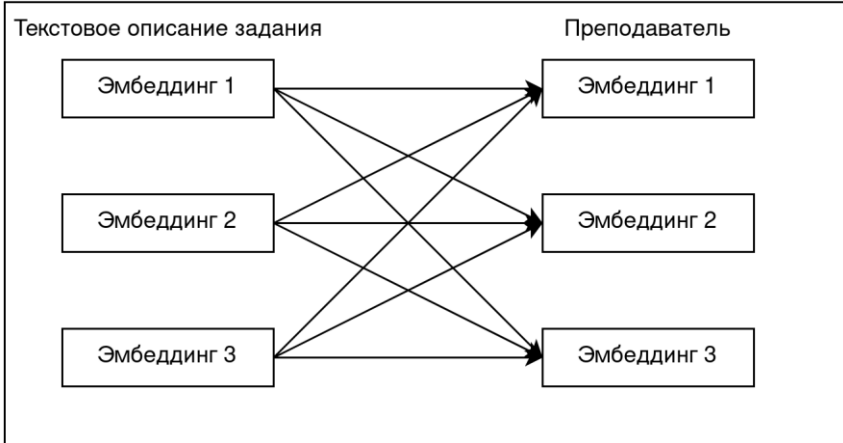
Описание алгоритма

На вход BERT'у подавались тексты ВКР. Потом она обрабатывала тексты и возвращала предложения в своем векторном пространстве. Потом эти предложения кластеризовывались и находилось обобщение текста.

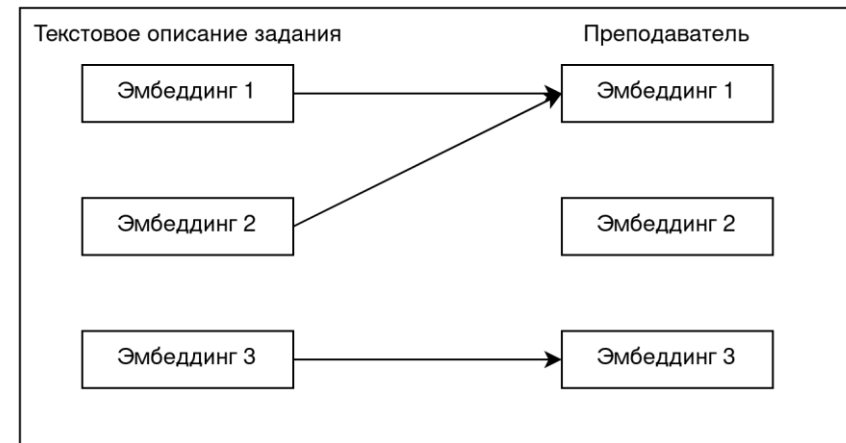
Input	[CLS]	my	dog	is	cute	[SEP]	he	likes	play	##ing	[SEP]
Token Embeddings	$E_{[CLS]}$	E_{my}	E_{dog}	E_{is}	E_{cute}	$E_{[SEP]}$	E_{he}	E_{likes}	E_{play}	$E_{\#ing}$	$E_{[SEP]}$
	+	+	+	+	+	+	+	+	+	+	+
Segment Embeddings	E_A	E_A	E_A	E_A	E_A	E_A	E_B	E_B	E_B	E_B	E_B
	+	+	+	+	+	+	+	+	+	+	+
Position Embeddings	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}



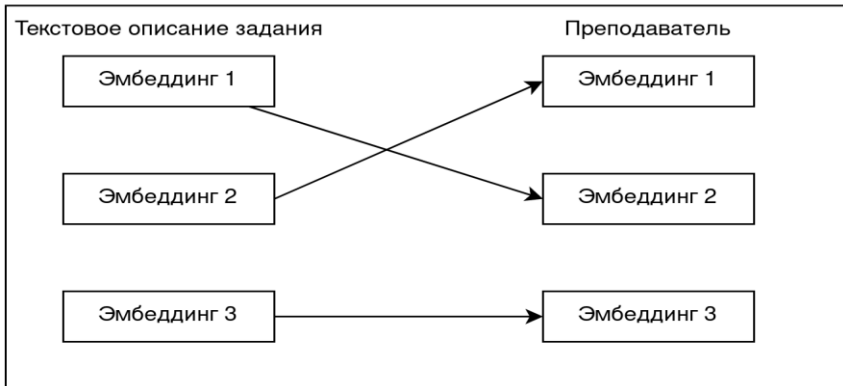
Подбор исполнителей



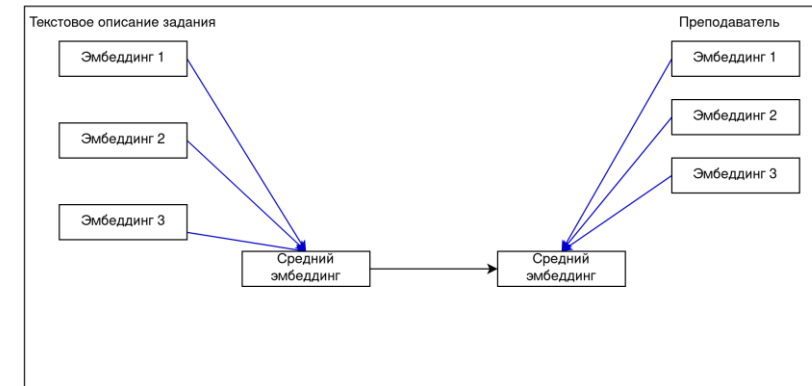
Все со всеми



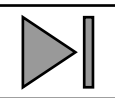
Минимальная разность



Минимальная разность
без повторов



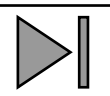
Усреднение эмбедингов





Расчет расстояний

- Модуль разницы элементов (Манхэттонновское расстояние)
- Квадрат разницы элементов (Евклидово расстояние)





Интерфейс

Введите текст:

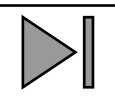
Выберите функцию расстояния

Эвклидово расстоян ▾

Выберите функцию подбора

Все со всеми ▾

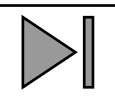
Готово



Тестирование

Для одного из тестов была выбрана выпускная квалификационная работа Абросимовой П. С. с темой «Разработка средств автоматизации расширения онтологии на основе данных интернет-источников» руководителем была Лядова Л. Н.

Тип алгоритма	Манхэттоновское расстояние	Расстояние Евклида
Все со всеми	Кушев В. О.	Кушев В. О.
Поиск минимального с повторами	Кычкин А. В	Кычкин А. В
Поиск минимального без повторов	Божья-Воля А. А.	Божья-Воля А. А.
Усреднение эмбеддингов	Кузнецов Д. Б.	Кузнецов Д. Б.



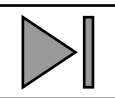


РЕЗУЛЬТАТ

Была разработана информационная система, целью которой было проверить гипотезу о том можно ли подобрать исполнителя под текстовое описание тендера.

На основании текстового описания был произведён поиск путём расчёта расстояния между подаваемым текстом и профилем сотрудников.

Было предложено несколько методов расчёта этого расстояния и было показано, что расстояние Эвклида показало лучшее качество работы на нескольких примерах.



Спасибо за внимание

Готов ответить на ваши вопросы



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

E-mail: risolomatin@edu.hse.ru

Соломатин Р. И.



Актуальность

Каждый день выкладывается много тендеров. В ВШЭ много людей с разными компетенциями, и надо по текстовому описанию тендера понять, кто его сможет сделать.



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9



Проблема и гипотеза

Проблема - автоматический поиск исполнителей под тендер, заданный текстовым описанием.

Гипотеза - можно ли автоматически подобрать исполнителя на основании анализа текстов ВКР?



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9



Существующие решения

- Много тендеров
- Много людей с разными компетенциями
- Тяжело масштабировать, потому что необходимо знать много про разных людей
- Тратится много времени



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9



Объект и предмет

Объект исследования - процесс поиска исполнителей по текстовому описанию тендера.

Предмет исследования - автоматизация процесса поиска исполнителей по текстовому описанию тендера.



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9



Проектирование



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9



Выбор технологий

Язык программирования	Библиотека для ML	Подготовка модели	Работа с сайтом	Работа с файлами	Интерфейс
Python	++	++	+	+	+
С#	+	+	+	+	-
Java	+	+	+	+	-
JavaScript	+	+	+	+	?



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9



База данных

Данные брались со сайта ВШЭ и помещались в базу данных



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9



Описание алгоритма

На вход BERT'у подавались тексты ВКР. Потом она обрабатывала тексты и возвращала предложения в своем векторном пространстве. Потом эти предложения кластеризовывались и находилось обобщение текста.



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9



Подбор исполнителей



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9



Расчет расстояний

- Модуль разницы элементов (Манхэттенское расстояние)
- Квадрат разницы элементов (Евклидово расстояние)



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9



Тестирование

Для одного из тестов была выбрана ситуационная квалификационная работа Абрамской П. С. с темой «Разработка средств автоматизации расширения онтологий на основе данных интернет-источников» руководителем была Людмила П. Н.

Тип алгоритма	Максимальное расстояние	Расстояние Евклида
Всё со всеми	Нудель В. О.	Нудель В. О.
Пресс минимального с повторами	Кликин А. В.	Кликин А. В.
Пресс минимального без повторов	Божин-Волк А. А.	Божин-Волк А. А.
Усреднение выборок	Кузнецов Д. Б.	Кузнецов Д. Б.



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9



РЕЗУЛЬТАТ

Была разработана информационная система, целью которой было проверить гипотезу о том можно ли подобрать исполнителя под текстовое описание тендера.

На основании текстового описания был произведен поиск путей расчёта расстояния между подаваемым текстом и профилем сотрудников.

Было предложено несколько методов расчёта этого расстояния и было показано, что расстояние Евклида показало лучшее качество работы на нескольких примерах.



21.01.2021

Соловьев И. И., «Понимание смысла текстовых описаний тендеров»

10/9