

“Nuclear IT Huck” МТС Лунк

**РАЗРАБОТКА СИСТЕМЫ НА ОСНОВЕ ИИ,
АНАЛИЗИРУЮЩЕЙ ПОЛЬЗОВАТЕЛЬСКИЕ
ОТВЕТЫ И ВОЗВРАЩАЮЩЕЙ ПОНЯТНОЕ
ОБЛАКО СЛОВ**

Дайте работу:

Дроздов Михаил, Павлов Дмитрий,
Дмитриевская Алиса, Махнев Кирилл

Getting data

Получение
данных из
google-таблицы
или .csv файла

Embeddings

Конвертация
текста в
эмбединги
(численное
представление
фраз)

Dimension reduction

Понижение
размерности
эмбедингов

Clustering

Решение задачи
кластеризации
сжатых
эмбедингов

Summarize

Выделение общей
мысли из фраз
каждого кластера

Visualization

Отображение
кластеров на
двумерной
плоскости

Получение данных

Google - форма

Имя Фамилия Отчество *

Мой ответ

Наименование Организации *

☐ МТС link

☐ МИФИ

☐ Другое

Что мотивирует вас работать больше? *

Мой ответ

Спасибо за ответ!

csv-файл

Отметка времени	Что мотивирует вас работать больше?	Имя Фамилия Отчч	Наименование Орг
26.09.2024 0:06:33	Премия	Махнев Кирилл Алекс	МИФИ
26.09.2024 0:06:33	Хорошая зарплата	Иванов Иван Иванович	МИФИ
26.09.2024 0:06:33	Друзья на работе	Петров Петр Петрович	МТС link
26.09.2024 0:06:33	Возможность карьерного роста	Смирнова Анна Серге	Другое
26.09.2024 0:06:33	Нормальная обстановка	Кузнецов Алексей Ник	МИФИ
26.09.2024 0:06:33	Да просто бабки нужны!	Васильева Светлана А	МТС link
26.09.2024 0:06:33	Получить опыт	Лебедев Сергей Викто	Другое
26.09.2024 0:06:33	Prosto zhal' ne uspet'	Сидорова Мария Влад	МИФИ
26.09.2024 0:06:33	Все ради команды	Григорьев Николай Юр	МТС link
26.09.2024 0:06:33	Крутые проекты!	Федорова Ольга Викто	Другое
26.09.2024 0:06:33	Заказов много, надо делать	Егорова Татьяна Дмит	МИФИ
26.09.2024 0:06:33	Работа ради удовольствия	Новиков Андрей Серге	МТС link
26.09.2024 0:06:33	Уважение коллег	Ковальчук Виктория А	Другое
26.09.2024 0:06:33	Хорошая команда	Зайцева Марина Анат	МИФИ
26.09.2024 0:06:33	Премии и бонусы	Семенов Игорь Валент	МТС link

Getting data

Получение
данных из
google-таблицы
или .csv файла

Embeddings

Конвертация
текста в
эмбединги
(численное
представление
фраз)

Dimension reduction

Понижение
размерности
эмбедингов

Clustering

Решение задачи
кластеризации
сжатых
эмбедингов

Summarize

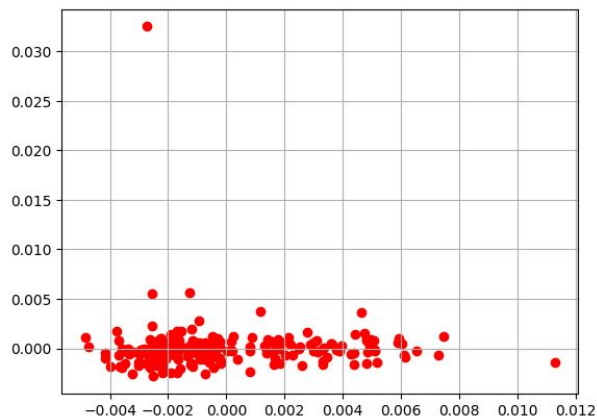
Выделение общей
мысли из фраз
каждого кластера

Visualization

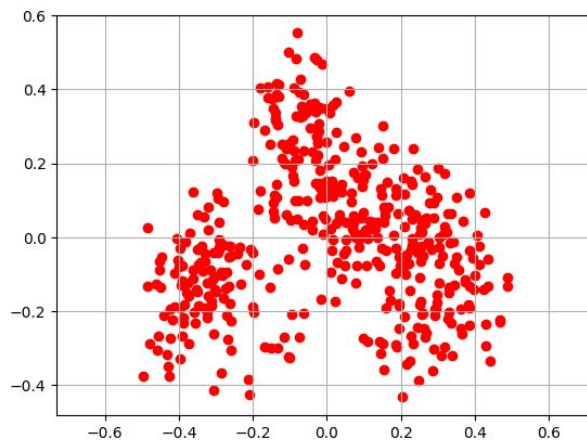
Отображение
кластеров на
двумерной
плоскости

Создание эмбеддингов фраз, понижение размерности

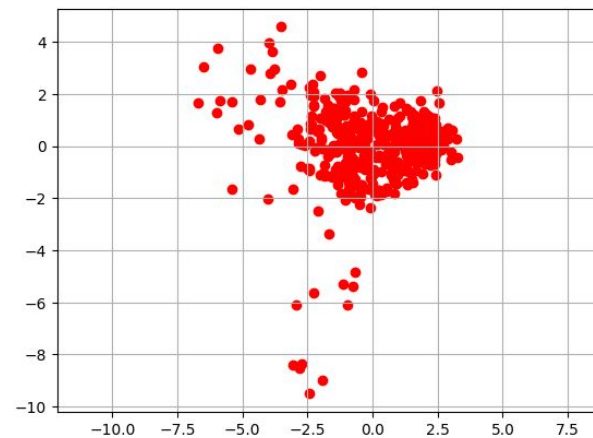
Word2Vec



USE



RuBERT



Getting data

Получение
данных из
google-таблицы
или .csv файла

Embeddings

Конвертация
текста в
эмбединги
(численное
представление
фраз)

Dimension reduction

Понижение
размерности
эмбедингов

Clustering

Решение задачи
кластеризации
сжатых
эмбедингов

Summarize

Выделение общей
мысли из фраз
каждого кластера

Visualization

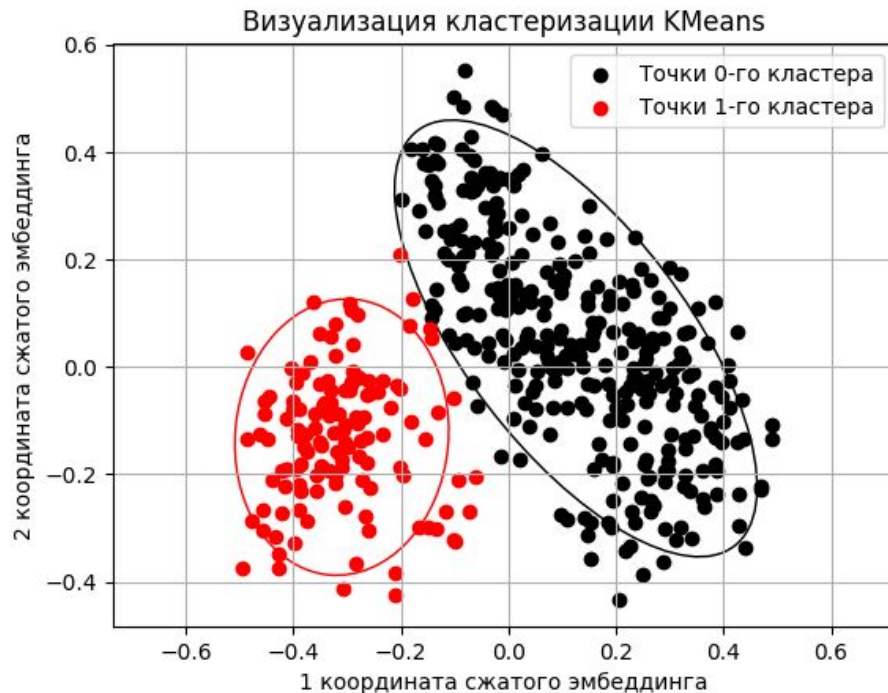
Отображение
кластеров на
двумерной
плоскости

Выбор модели кластеризации

Опробованы методы:
Birch, DBSCAN, KMeans

Выбран KMeans:

- Прост в реализации
- Эффективен



Getting data

Получение
данных из
google-таблицы
или .csv файла

Embeddings

Конвертация
текста в
эмбединги
(численное
представление
фраз)

Dimension reduction

Понижение
размерности
эмбедингов

Clustering

Решение задачи
кластеризации
сжатых
эмбедингов

Summarize

Выделение общей
мысли из фраз
каждого кластера

Visualization

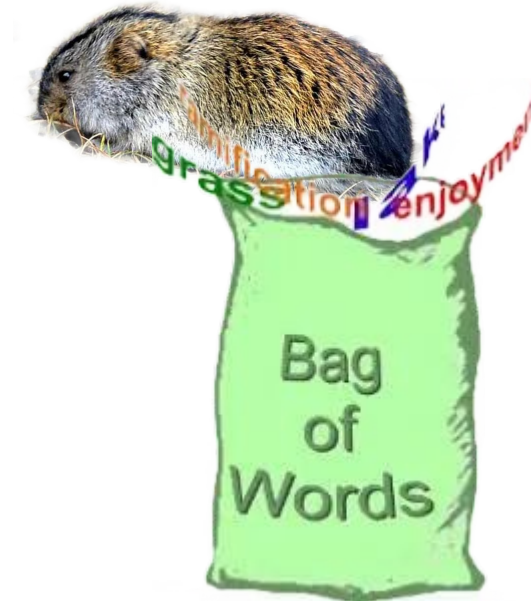
Отображение
кластеров на
двумерной
плоскости

Выделение главной мысли в кластере

Использование API ChatGPT
для суммаризации



Леммингз + мешок слов



Getting data

Получение
данных из
google-таблицы
или .csv файла

Embeddings

Конвертация
текста в
эмбединги
(численное
представление
фраз)

Dimension reduction

Понижение
размерности
эмбедингов

Clustering

Решение задачи
кластеризации
сжатых
эмбедингов

Summarize

Выделение общей
мысли из фраз
каждого кластера

Visualization

Отображение
кластеров на
двумерной
плоскости



СПАСИБО ЗА ВНИМАНИЕ!

Дайте работу:

Дроздов Михаил, Павлов Дмитрий,
Дмитриевская Алиса, Махнев Кирилл

Анализ ответов

Загружайте и анализируйте данные, чтобы получить информацию о мнениях, предпочтениях и мотивах

Данная программа использует искусственный интеллект для обработки текстовых данных. Конечным результатом работы алгоритма является построение на двумерной плоскости смысловых кластеров, содержащих близкие по смыслу выражения. Название кластеров отражает основную мысль выражений, попавших в этот кластер. Для более подробного ознакомления с алгоритмом работы рекомендуем ознакомиться с файлом README.md в репозитории

Руководство пользователя:

- 1) Выберите файл с данными, отвечающими вашему опросу в google forms
- 2) Выберите модель (RuBert или USE) для образования эмбедингов
- 3) Нажмите кнопку "Запустить модель"

Если все прошло успешно, вы увидите, что на графике справа появилось изображение кластеров. При нажатии курсором на кластер, под графиком выведутся фразы, попавшие в этот кластер

Выберите модель для создания эмбедингов

Загружена модель RuBert



Загрузить RuBERT



Загрузить Universal sentence encoder

Выберите файл для извлечения данных

Выбранный файл: /home/drozdovmk/Projects/NLP_hakaton/Hakaton-mts-link/data/motivations.csv



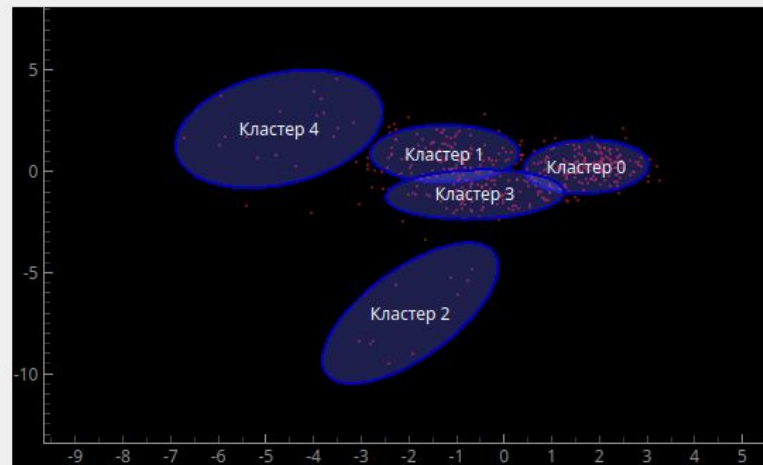
Получить данные из Google forms



Получить данные из csv

☒ Использовать API ChatGPT для выделения главной мысли в кластере

Запустить модель



Описание кластера 0:

Количество объектов в кластере: 189

Главная мысль: Поддержка и признание мотивируют.

Когда моя работа приносит пользу другим — это лучшая мотивация.

Увлекательный процесс работы сам по себе мотивирует.

Работаю, потому что сам процесс приносит радость.

Мотивирует возможность всегда быть на плаву.

Ну, чем выше зарплата, тем больше мотивации, это логично.

Хочу стать лучшей версией себя.

Амбиции подталкивают меня работать больше.

Когда уважаешь людей, с которыми работаешь, появляется мотивация работать ещё

Интересная работа сама по себе — лучшая мотивация.

Атмосфера взаимопомощи сильно мотивирует.

Достижение целей наполняет энергией.

Знание, что у меня стабильная работа, даёт мне уверенность.

Когда есть с кем поделиться победами и трудностями — это мотивирует.

Работаю ради того, чтобы принести пользу окружающим.