

Мой первый чат-бот «Искусственный риелтор».

Подтема

Прототип чат-бота для продажи моего дома.





Дмитрий Яковлев

Закончил курсы повышения квалификации в МГТУ им. Баумана.

Закончил факультет искусственного интеллекта GeekBrains..

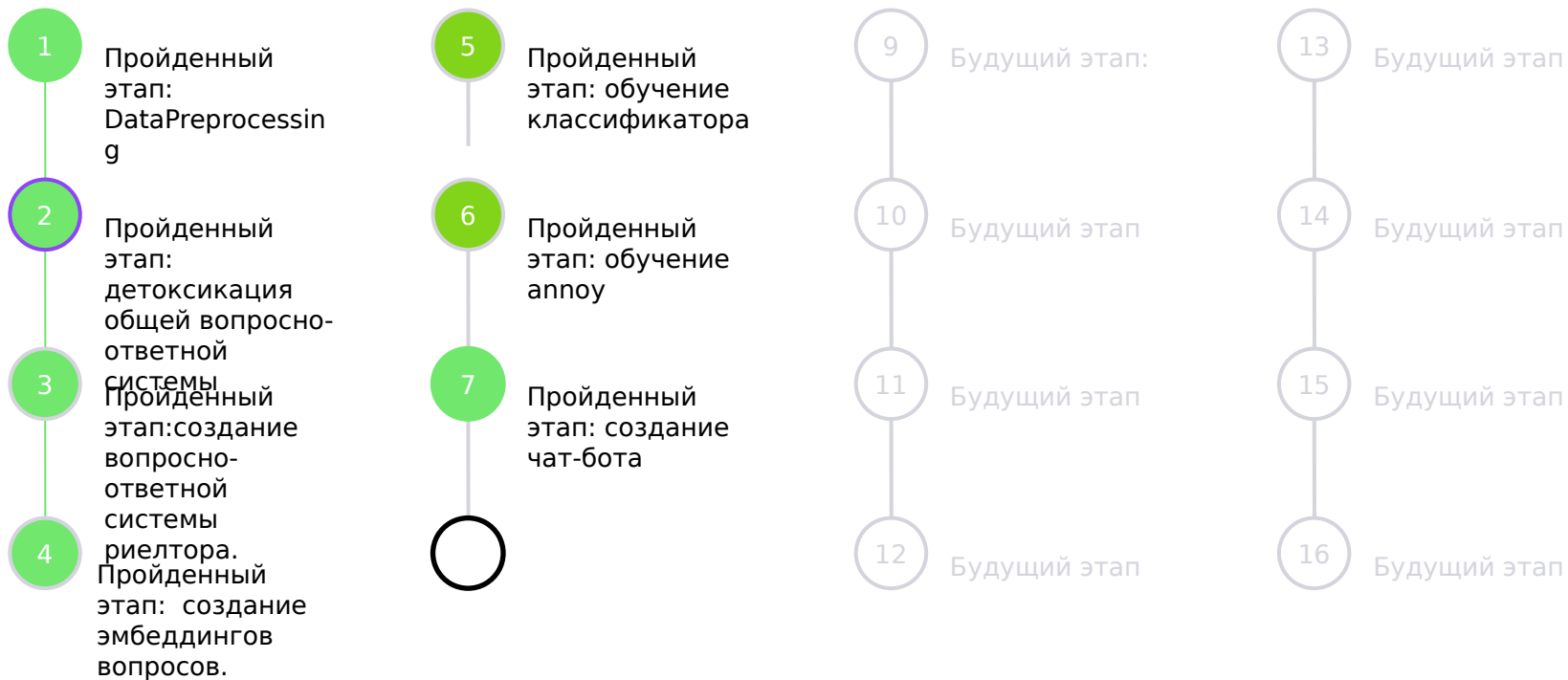
Закончил Сибирскую аэрокосмическую академию по специальности «Системы управления ракетно-космическими объектами и комплексами летательных аппаратов»

Немного о себе.

Инженер по работе с проектировщиками и клиентами на предприятии, производящем высоковольтное оборудование.



План проекта





Решение задачи / План работы

- Исследование вопросно-ответной системы от mail.ru.
- Создание вопросно-ответной системы «риелтора».
- Чистка и выборка по вопросно-ответной системе от mail.ru.
- Создание бертовых эмбедингов.
- Разработка классификатора.
- Создание вопросно-ответного «движка» от annou.
- Создание чат-бота.



Трудности

Поскольку у чат-бота коммерческая направленность, а до ChatGPT мне далеко, с генерацией ответов не стоило связываться. Трудности возникли в создании вопросно-ответной системы «риелтора». Это большой объем работы и я с ним до конца справиться не успел. Трудно было почистить и сделать выборку из вопросно-ответной системы от mail.ru. Много было мусорных и токсичных сообщений.

К сожалению бертовые эмбединги из-за огромного размера не возможно закатать на Github. Также не закатал файл setting.py с секретными ключами. Поэтому запустить чат-бота в таком состоянии не получится.

С 2022 года MongoDB не работает с Россией, поэтому MongoDB пришлось отключить.



Выбор классификатора

```
altaypower
def choice_data(question, morpher=morpher, sw=sw, exclude=exclude,
                morpher_qwest_tokens=morpher_qwest_tokens):
    spls = preprocess_txt(question, morpher, sw, exclude)
    cnt=0
    for i in spls:
        if i in morpher_qwest_tokens:
            cnt+=1
    question = ' '.join(spls)
    return cnt, question
```

Поскольку вопросно-ответная система от mail.ru во много порядков больше чем вопросно-ответная система «риелтора», от обучения ML-классификатора для этого дела пришлось отказаться. Настроил простейший классификатор: если в вопросе присутствуют слова из вопросов для «риелтора», значит этот вопрос про покупку дома. Иначе это общий вопрос и ответ должен быть из вопросно-ответной системы от mail.ru.



Чат-бот

В чат-боте очень много функций, которые очень долго здесь расписывать. В целом работу чат-бота хорошо иллюстрируют следующие картинки:

```
Run: bot x
/home/jb/PycharmProjects/Repo_NLP_Chat_Bot_MongoDB/venv/bin/python /home/jb/PycharmProjects/Repo_NLP_Chat_Bot_MongoDB/bot.py
[nltk_data] Downloading package stopwords to /home/jb/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to /home/jb/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
2023-02-09 18:44:21.421381: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcudart.so.11.0.3': dlopen(/home/jb/nltk_data/.../libcudart.so.11.0.3, 1): libcuda.so.1: cannot open shared object file: No such file or directory
2023-02-09 18:44:21.421406: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cudart dlerror if you do not have a GPU installed
Downloading config.json: 100%|██████████| 611/611 [00:00<00:00, 194kB/s]
Downloading tf_model.h5: 100%|██████████| 1.756/1.756 [02:59<00:00, 10.5MB/s]
2023-02-09 18:47:26.742898: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcudart.so.11.0.3': dlopen(/home/jb/nltk_data/.../libcudart.so.11.0.3, 1): libcuda.so.1: cannot open shared object file: No such file or directory
2023-02-09 18:47:26.742927: W tensorflow/stream_executor/cuda/cuda_driver.cc:269] failed call to cuInit: UNKNOWN ERROR (303)
2023-02-09 18:47:26.742949: I tensorflow/stream_executor/cuda/cuda_diagnostics.cc:156] kernel driver does not appear to be running on this GPU (UID 0).
2023-02-09 18:47:26.767484: I tensorflow/core/platform/cpu_feature_guard.cc:193] This TensorFlow binary is optimized with oneAPI Deep Neural Compiler (oneDNN) for Linux.
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
2023-02-09 18:47:27.183916: W tensorflow/core/framework/cpu_allocator_impl.cc:82] Allocation of 1539542016 exceeds 10% of free system memory
2023-02-09 18:47:31.436370: W tensorflow/core/framework/cpu_allocator_impl.cc:82] Allocation of 1539542016 exceeds 10% of free system memory
2023-02-09 18:47:31.990981: W tensorflow/core/framework/cpu_allocator_impl.cc:82] Allocation of 1539542016 exceeds 10% of free system memory
2023-02-09 18:47:36.544862: W tensorflow/core/framework/cpu_allocator_impl.cc:82] Allocation of 1539542016 exceeds 10% of free system memory
2023-02-09 18:47:37.743804: W tensorflow/core/framework/cpu_allocator_impl.cc:82] Allocation of 1539542016 exceeds 10% of free system memory
All model checkpoint layers were used when initializing TFBertModel.
```



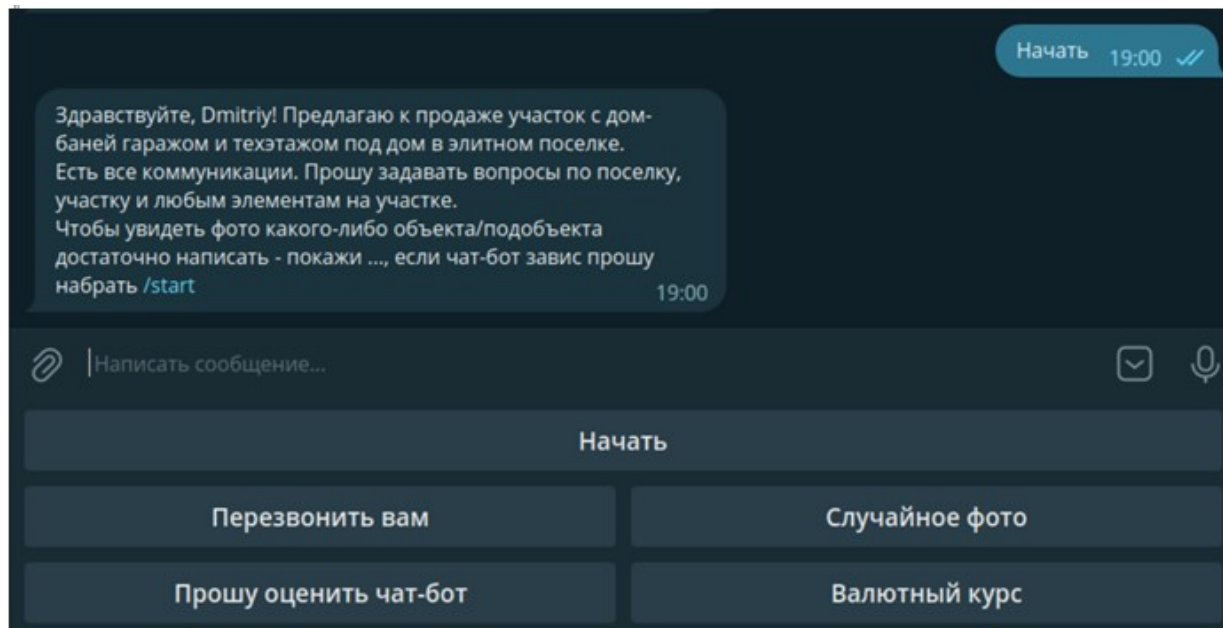
Чат-бот

```
Repo_NLP_Chat_Bot_MongoDB [Re] Package requirements 'absl-py==1.3.0', 'cachetools==5.2.0', 'catboost==1.... Install requirements Ignore requirements
> database
Run: bot x
2023-02-09 18:47:31.990981: W tensorflow/core/framework/cpu_allocator_impl.cc:82] Allocation of 1539542016 exceeds 10% of free system memory
2023-02-09 18:47:36.544862: W tensorflow/core/framework/cpu_allocator_impl.cc:82] Allocation of 1539542016 exceeds 10% of free system memory
2023-02-09 18:47:37.743804: W tensorflow/core/framework/cpu_allocator_impl.cc:82] Allocation of 1539542016 exceeds 10% of free system memory
All model checkpoint layers were used when initializing TFBertModel.

All the layers of TFBertModel were initialized from the model checkpoint at setu4993/LaBSE.
If your task is similar to the task the model of the checkpoint was trained on, you can already use TFBertModel for predictions with
Downloading tokenizer_config.json: 100%|██████████| 327/327 [00:00<00:00, 128kB/s]
Downloading tokenizer.json: 100%|██████████| 13.0M/13.0M [00:02<00:00, 6.63MB/s]
Downloading special_tokens_map.json: 100%|██████████| 125/125 [00:00<00:00, 48.5kB/s]
Кто-то отправил команду /start.
покажи баню
как проехать
есть газ
есть газ?
какие коммуникации?
есть канализация?
есть водопровод?
как проехать?
```

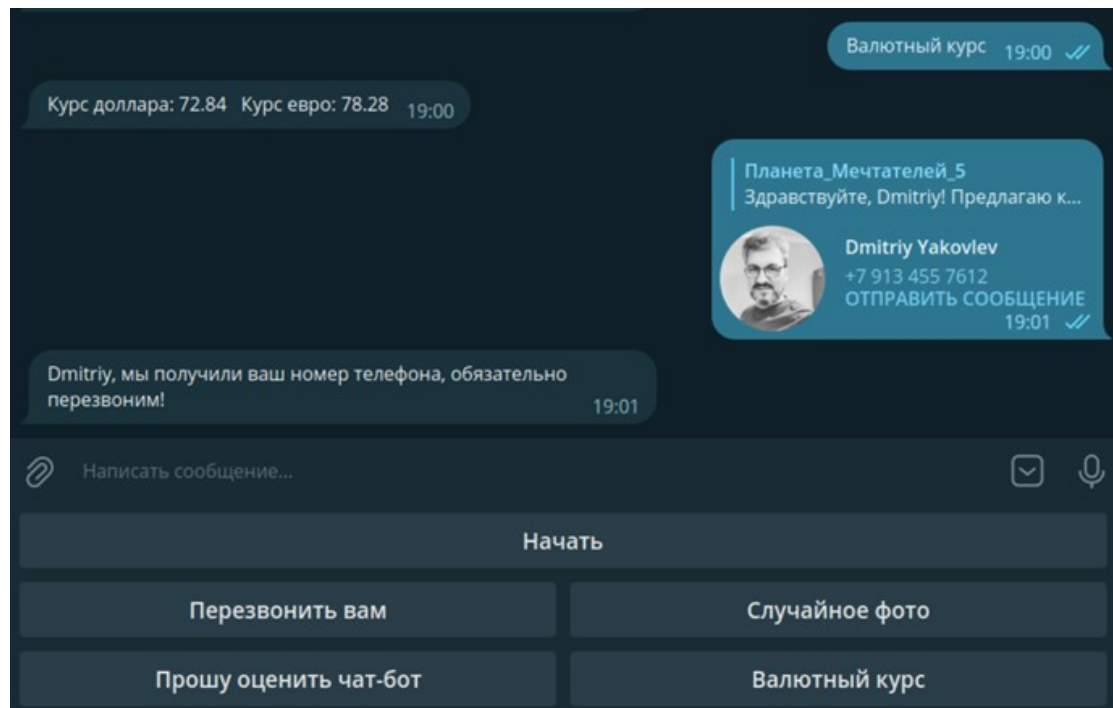



Чат-бот



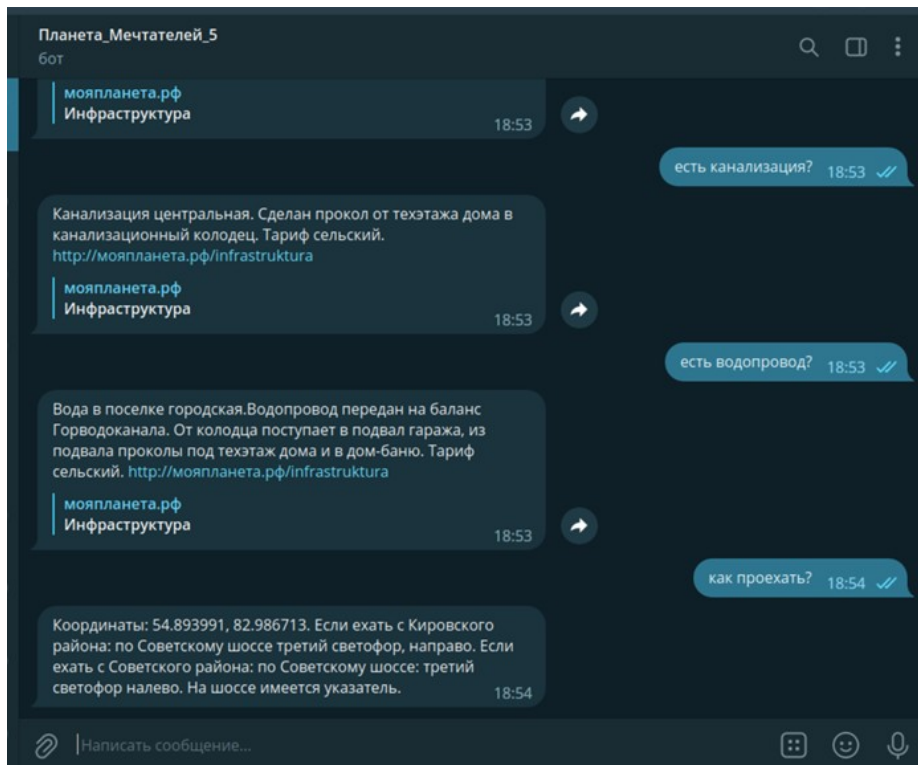


Чат-бот



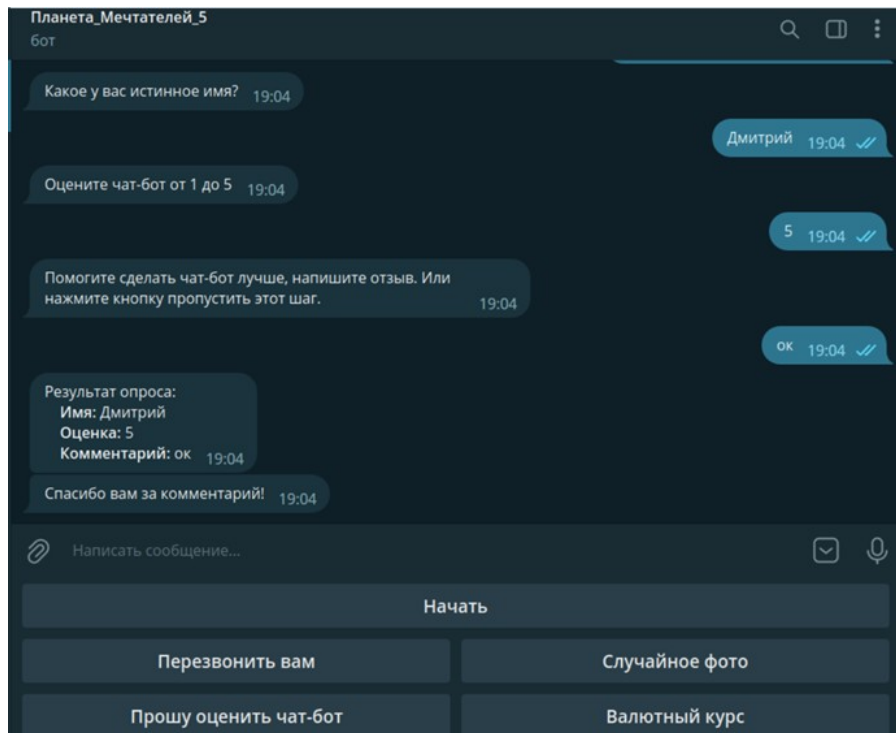


Чат-бот





Чат-бот





Предложения по проекту

В первую очередь дополнить вопросно-ответную систему «риелтора». Убрать кнопки валютного курса и случайного фото и сделать кнопку “FAQ” с общим обзором усадьбы и обучающим пользованию ботом видео. Создать веб-интерфейс «риелтора». Настроить работу базы данных.