SYNCRETIS

ТЕХНИЧЕСКОЕ ЗАДАНИЕ

на разработку

моделей для голосового офисного помощника.

SYNCRETIS

Содержание.

SYNCRETIS

1 Цель.

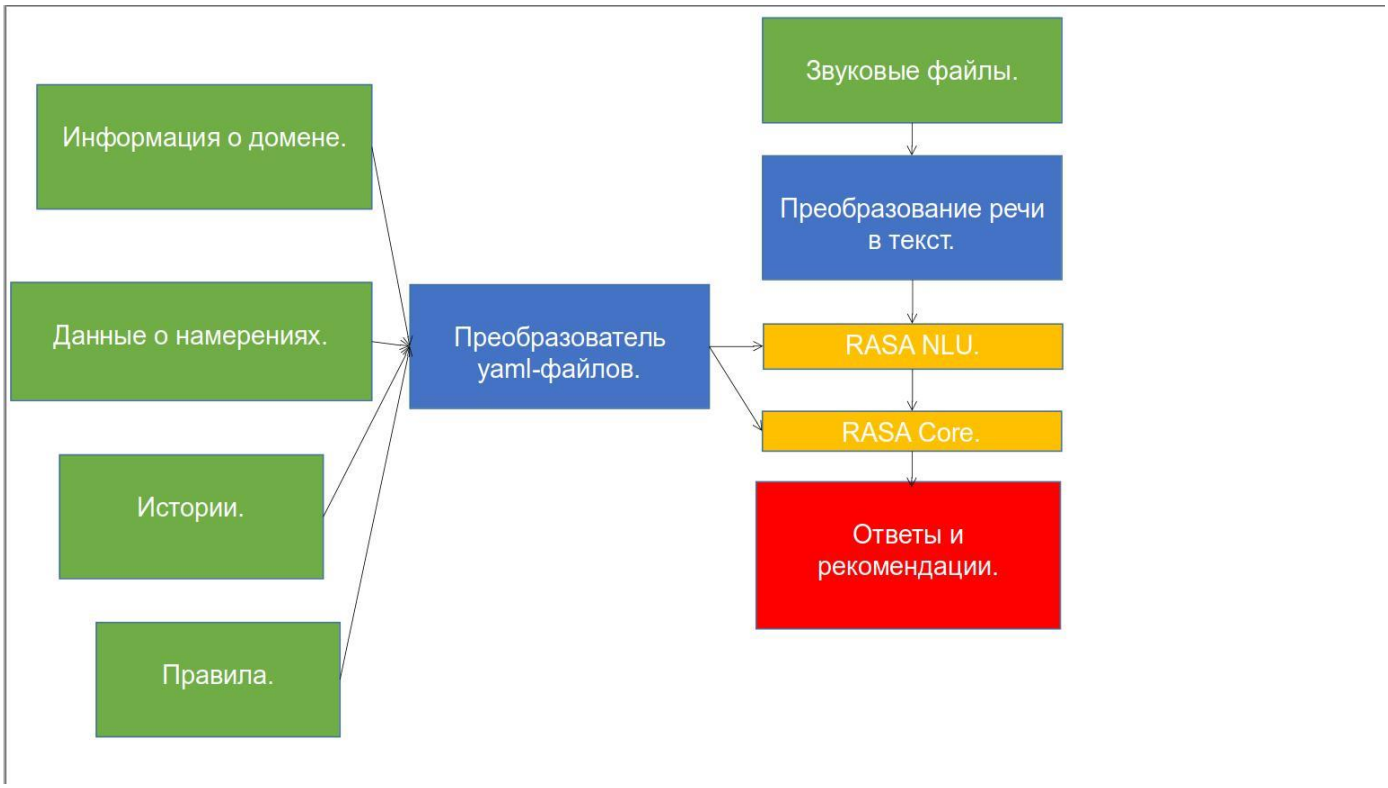
Нашей целью является создание моделей машинного обучения для голосового офисного помощника.

С точки зрения пользователя функционал выглядит следующим образом: пользователь произносит определённые фразы и видит на экране ответы модели. Речь идёт о специализированном чат-боте (офисная тематика). В финальной версии должна быть возможность добавления пользовательских интентов. В первоначальной версии нужна только поддержка русского языка, а в финальной версии предполагается поддержка английского языка.

Предполагается сделать модель по превращению речи в текст и сделать чат бот на основе RASA (фреймворк для создания чат ботов, широко использующийся в индустрии). Отметим, что у RASA есть и open source версия, и платная версия. Предполагается использовать open source версию.

SYNCRETIS

2 Схема решения.



Зелёное --- входные данные.

Синее и жёлтое --- блоки решения.

Красное --- выходные данные.

SYNCRETIS

3 Входные данные.

Опишем здесь входные данные:

- **Звуковые файлы.** Так как модель принимает на вход речь, естественным способом её хранения являются звуковые файлы.
- **Информация о домене.** Домен --- это большая область, в которой работает специализированный чатбот. У чатбота може быть, как один домен, так и несколько доменов. Доменами могут быть разные офисы (теоретически у разных офисов может быть разная структура общения, разные реплики).
- **Данные о намерениях.** Намерения или интенты --- это то, что хочет пользователь. В финальной версии можно будет добавлять интенты. Пример интента: сообщить свои ФИО.
- **Истории.** Истории --- это сценарии развития диалога. Пример истории: заявление на отпуск.
- **Правила.** Правила --- это сценарии поведения. Например, говорить “Пока”, если у пользователя интент “Попрощаться”. Разница между историями и правилами в том, что для историй характерны множество путей развития.

Эти входные данные предполагается передавать модели в реальном времени в ходе её работы.

Модели также понадобятся данные на этапе обучения.

4 Модель распознавания речи.

Предполагается использовать предобученные модели (ESPnet2, kaldi, XLSR-53 и др.) и дообучать их на данных, предоставляемых заказчиком.

Данная модель принимает на вход звуковой файл и возвращает текст, который предоставляет обычному чат-боту. В этом тексте могут быть ошибки, опечатки, со всем этим должен справиться обычный чат-бот на основе RASA.

Вызовы модели логгируются для возможного улучшения её работы в дальнейшем.

SYNCRETIS

5 Преобразователь yaml-файлов.

RASA принимает на вход большое количество yaml-файлов. Однако они содержат большое количество служебной информации (NLP-модели, которые надо использовать внутри, например). Не разумно всю эту информацию объявлять входными данными модели. Поэтому есть смысл объявить входными данными, только небольшую, наиболее релевантную часть, а остальное заполнять в преобразователе yaml-файлов.

6 RASA NLU.

Это часть, отвечающая за понимание пользователя. Блок принимает на вход текстовую реплику пользователя. Далее исправляются опечатки, определяются намерения пользователя с помощью различных NLP библиотек. Предпочтение предполагается отдавать spaCy (поддерживается и английский, и русский языки, библиотека широко используется в индустрии).

Предполагается использовать предобученные модели выделения сущностей и дообучать их на данных, предоставляемых заказчиком.

Вызовы модели логгируются для возможного улучшения работы в дальнейшем.

7 RASA Core.

Эта часть RASA отвечает за управление диалогом и за генерацию ответов (бывает, что этот блок разделяют на 2 подблока).

Надо предусмотреть возможность переспрашивать пользователя для улучшения качества работы. Надо предусмотреть специальный класс интентов для того, чтобы реагировать на нерелевантные ответы.

В финальной версии можно предусмотреть пути для разных видов пользователей (довольного и недовольного пользователей).

Нам понадобятся данные от заказчика для определения различных сценариев развития диалога.

Вызовы модели логгируются для возможного улучшения работы в дальнейшем.

SYNCRETIS

8 Выходные данные.

На выходе модель должна предоставлять выходные реплику пользователю в текстовом формате. При желании можно использовать предобученную нейронную сеть для превращения текста в речь.

Кроме того, модель должна выдавать определённые действия (условные строки-команды для родительского приложения, с которым предстоит интегрироваться).

SYNCRETIS

9 MVP и дальнейшая работа.

MVP рассчитан на 3 месяца, предполагается, что его будет делать 1 человек в режиме full time. Предполагается получение данных от заказчика для дообучения моделей.

Предполагается сделать следующее:

- Модель превращения голоса в текст с учётом данных заказчика.
- Чат-бот на основе RASA с учётом данных заказчика.
- Входные и выходные данные модели показывают возможную интеграцию с приложением заказчика.

Дальнейшая работа:

- Улучшение моделей в соотвествии с пожеланиями заказчика.
- Поддержка английского языка.
- Поддержка разных типов пользователей.
- Превращение текстовых реплик модели в речь.