

УДК 004.89

DOI: [10.26102/2310-6018/2020.29.2.041](https://doi.org/10.26102/2310-6018/2020.29.2.041)

Целевой чат-бот на основе машинного обучения

Т.М.Ч. Нгуен, М.В. Щербаков

*Волгоградский государственный технический университет,
Волгоград, Российская Федерация*

Резюме: В настоящее время чат-боты становятся очень популярными во многих областях, таких как бизнес, банковское дело, здравоохранение, учеба, путешествия и т.д. Популярность платформ обмена сообщениями, таких как Telegram, Messenger, Whatsapp и других сделала чат-ботов не только популярными, но и стало тенденцией в будущем. С конца декабря 2019 года из-за начала пандемии COVID-19 наступил серьезный кризис в области здравоохранения во всем мире. В связи с этим крайне важно предоставлять информацию об эпидемии всем людям. Многие правительства и организации запустили чат-ботов для информирования населения о COVID-19. Однако эти чат-боты ограничены, так как основаны на правилах, и понимают ограниченный набор вопросов, введенных пользователем. Таким образом, создание чат-бота на основе машинного обучения для информации о коронавирусе представляет собой актуальную задачу. Целью исследования является разработка чат-бота для поиска информации про коронавирусную инфекцию COVID-19. Описан метод проектирования и разработка чат-бота на фреймворке RASA, а также испытания разработанного прототипа. Созданы три модели чат-бота: базовая модель (B), базовая модель с синонимами (BS), базовая модель с синонимами и шумами (BSS). Проведена оценка эффективности трех моделей по показателям: доля, точность и F-мера. Результаты анализа показали, что модели BS и BSS лучше, чем модель B.

Ключевые слова: чат-бот, обработка естественного языка, бессерверная архитектура, намерение, сущности, RASA, COVID-19.

Для цитирования: Нгуен Т.М.Ч., Щербаков М.В. Целевой чат-бот на основе машинного обучения. *Моделирование, оптимизация и информационные технологии*. 2020;8(2). Доступно по: https://moit.vivt.ru/wp-content/uploads/2020/05/NguyenShcherbakov_2_20_1.pdf DOI: 10.26102/2310-6018/2020.29.2.041

A GOAL-ORIENTED CHATBOT BASED ON MACHINE LEARNING

T.M.T Nguyen, M.V. Shcherbakov

Volgograd State Technical University, Volgograd, Russian Federation

Abstract: Nowadays chatbots are becoming very popular in many areas, such as business, banking, healthcare, study, travel tips, etc. The popularity of messaging platforms such as Telegram, Messenger, Whatsapp, and others has made chatbots not only popular but also become a trend in the future. Since the end of December 2019, the onset of the COVID-19 pandemic has brought about a major global health crisis. Therefore, it is extremely important to provide information about the epidemic to all people. Many governments and organizations have launched chatbots to inform the public about COVID-19. However, these chat rules are limited as they understand a limited set of questions entered by users. Thereby, creating a chatbot based on machine learning for coronavirus information is an urgent task. The purpose of the study is the development of a chatbot for searching for information about COVID-19 coronavirus infection. The method of designing and developing a chatbot on the RASA framework, as well as testing of the developed prototype, are described. Three chatbot models were created: the baseline model (B), the baseline model with synonyms (BS), and the baseline model with

synonyms and noises (BSS). The effectiveness of three models was evaluated based on the following indicators: accuracy, precision, and F-measure. The analysis results showed that the BS and BSS models are better than the B model.

Keywords: chat bot, natural language processing, serverless, intent, entities, RASA, COVID-19.

For citation: Nguyen T.M.T., Shcherbakov M.V. *A goal-oriented chatbot based on machine learning. Modeling, optimization and information technology.* 2020;8(2). Available from: https://moit.vivt.ru/wp-content/uploads/2020/05/NguyenShcherbakov_2_20_1.pdf DOI: 10.26102/2310-6018/2020.29.2.041 (In Russ).

Введение

В настоящее время в области искусственного интеллекта повысилось внимание людей к чат-ботам. Чат-бот – это компьютерная программа, способная взаимодействовать с пользователем с помощью текстовых или аудио сообщений. Сейчас использование чат-ботов в деятельности организаций и предприятий становится все более популярным. Чат-боты могут подобрать лучший рейс, забронировать гостиницу, выбрать покупку, то есть они представляют собой новую подотрасль обслуживания и ассистирования. Использование чат-бота дает много преимуществ для компаний и организаций [1] всегда работающих 24/24 в течение 365 дней без отдыха. С его помощью они могут общаться с тысячами клиентов одновременно (клиенты получают ответы немедленно, без ожидания), автоматически отправлять сообщения о работе с клиентами, автоматически собирать информацию и отчеты. В здравоохранении чат-боты помогают быстро и просто осуществлять запись к врачу, давать советы пациентам. Поэтому чат-боты становятся все более популярными, и все больше пользователей хотят с ними общаться. Благодаря вышеперечисленным преимуществам чат-боты открывают будущее в области здравоохранения, экономике и в других областях.

В последнее время многие веб-сервисы предоставляют разработчикам программные интерфейсы для разработки чат-ботов, интегрированных в этот сервис. Такие интерфейсы предоставляют социальные сети Facebook, Twitter и ВКонтакте, а также мессенджеры WhatsApp, Telegram.

Обработка естественного языка (англ. Natural Language Processing, NLP), как ядро взаимодействия в чате, встроено в облачный когнитивный сервис. Есть много систем, реализующих подходы искусственного интеллекта, таких как IBM Watson, Wit.AI, Dialog Flow и Rasa. Rasa Stack [2] - это наборы инструментов NLP с открытым исходным кодом, ориентированные в первую очередь на создание чат-ботов. Это один из самых эффективных и экономичных по времени инструментов для создания сложных чат-ботов.

Бессерверная архитектура рассматривается как альтернативный способ создания серверных приложений. В [3-5] было представлено построение чат-бота с использованием бессерверной модели.

Постановка задачи

Коронавирусная инфекция COVID-19 [6] продолжает широко и быстро распространяться в мире и в России. Пандемия COVID-19 серьезно повлияла на ситуацию во всем мире. Во многих странах эпидемия коронавируса шла на спад в мае и начале июня. Помимо этого, эксперты говорят, что эпидемия коронавируса может повториться осенью [7]. Все больше и больше людей болеют, а вакцины еще нет, поэтому люди ищут информацию о болезни ежедневно. Чтобы узнать информацию о

коронавирусе пользователи чаще всего заходят в интернет или социальные сети. Поэтому было принято решение разработать чат-бота для помощи поиска информации о коронавирусной инфекции.

Метод

В процессе исследования был создан прототип чат-бота для мессенджера Slack, написанном на языке программирования Python 3.6 с использованием платформа Rasa, для взаимодействия с Slack через HTTP запросы. На Рисунке 1 показан схема работы системы и поток данных-ответов. Система состоит из трех чатсей: Webhook [8], платформа Rasa, API для получения данных .

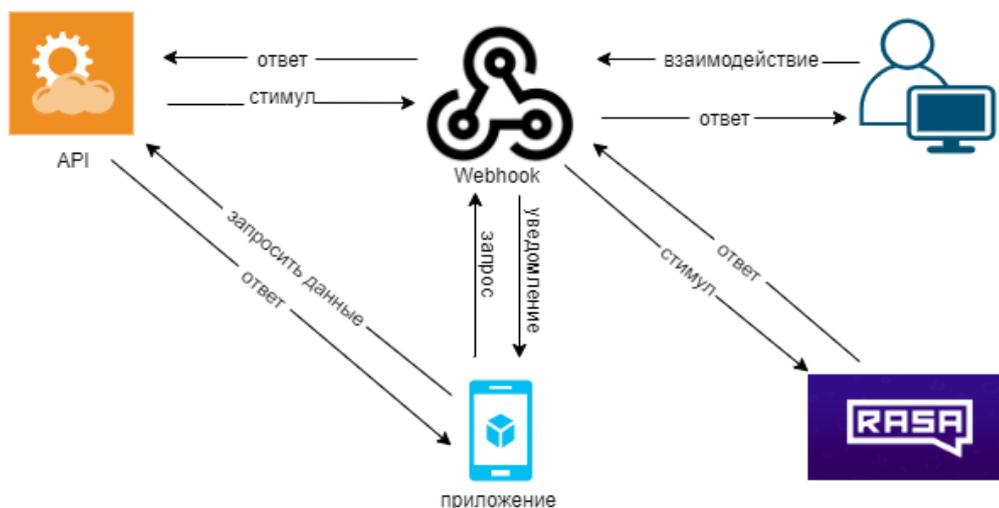


Рисунок 1 - Бессерверная архитектура чат-бота с использованием платформа Rasa
 Figure 1 - The serverless chatbot architecture using the Rasa platform

Webhook - механизм оповещения системы о событиях. Чат-бот использует технологию Webhook для аутентификации и для передачи уведомлений о происходящих событиях вашему приложению. С точки зрения программирования всё это сводится к работе функций обратного вызова для обработки HTTP-запросов, которые получают данные о событиях, типа полученных чат-ботом сообщений.

RASA состоит из двух частей: Rasa NLU и Rasa CORE.

Rasa NLU. Это библиотека высокого уровня, аналогичная онлайн-сервисам, которая анализирует пользовательские сообщения в диалогах, состоящих из интенгов (намерений) и набора соответствующих сущностей. Rasa NLU [9] следует модульной конструкции, опираясь на существующие библиотеки обработки естественного языка (NLP) и машинного обучения, такие как spaCy, scikit-learn, sklearn-crfsuite. Конвейер Rasa NLU состоит из процесса токенизации, за которым следует тегирование частей речи (POS) с помощью библиотеки spaCy NLP.

Для обучения Rasa NLU список вопросов, помеченных интенгов и сущностями, подготовлен с использованием читаемого человеком формата JSON или markdown [10]. В данной статье в чат-боте включены 14 интенгов.

Rasa CORE. Как указано выше Rasa NLU извлекает сущности и интенги, а Rasa Core [11] служит для подготовки ответов пользователю, управляет потоком разговоров, высказываниями, действиями. Этот раздел становится очень важным, когда у нас есть различные интенги в чат-боте и их последующие вопросы или ответы.

COVID-19 API. Для получения статистики по заболеваемости в мире и во всех странах мы использовали Covid-19 API [12]. API предоставляют данные в формате JSON.

Разговорный поток. Разговорный поток - это поток общения, существующий в службах NLP. Поток беседы направляет разговор, поэтому он может проходить в соответствии с определенными правилами. Пользователь может что-то спросить, но, чтобы достичь функциональности чат-бота, пользователь должен что-то отправить. В качестве эталона разговорного потока существует некоторое готовое приложение, такое как Covi19bot, которое может дать базовую структуру разговора. Пример разговорного потока чат-бота, который был создан Rasa Core, показан на Рисунке 2.

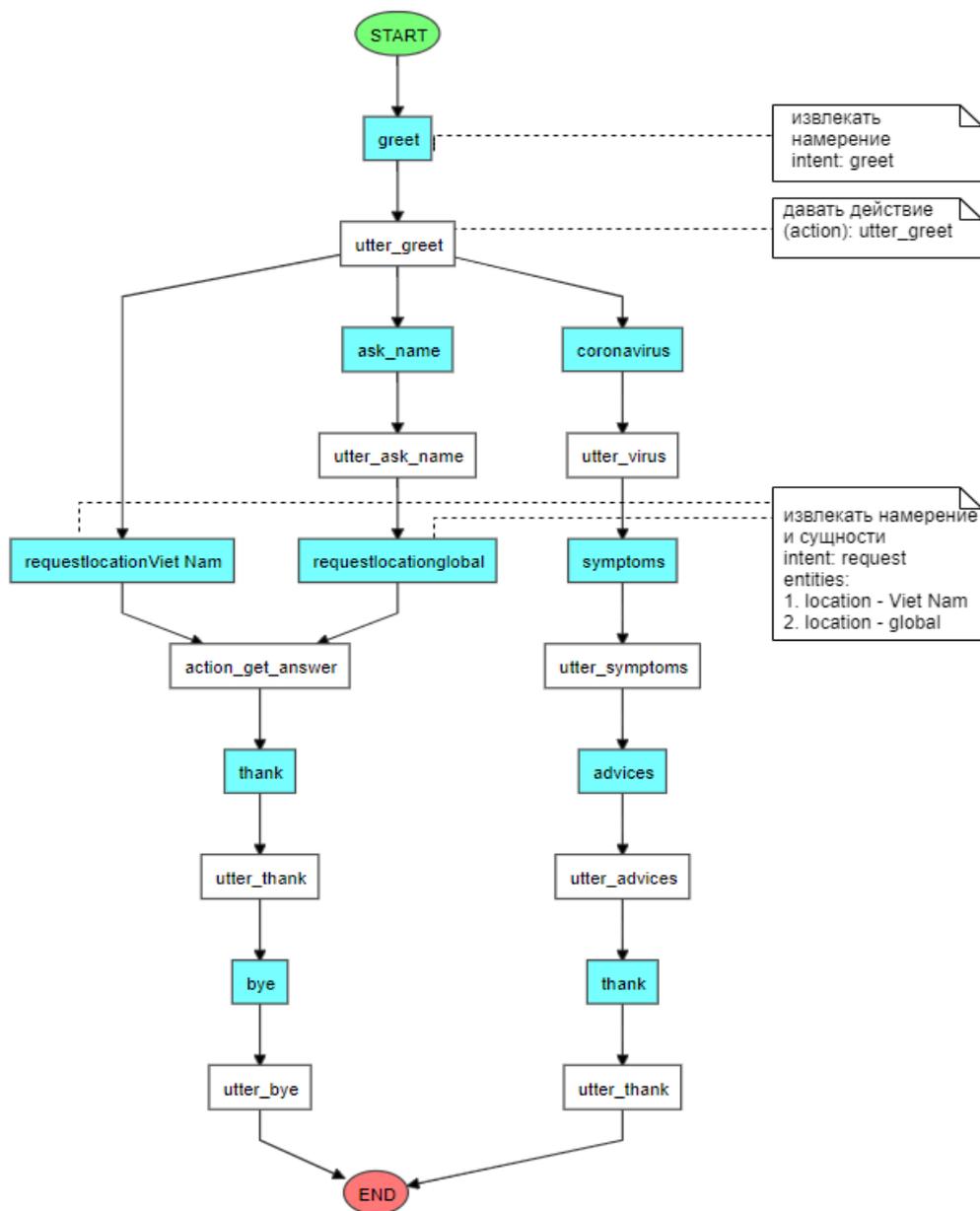


Рисунок 2 - Пример разговорного потока чат-бота
Figure 2 - An example of a chatbot conversation flow

Эксперимент и результаты

Rasa предоставляет два основных метода для построения обучающих данных для ботов:

- предварительно обученные интенеры (Intent_classifier_sklearn): классификация интенеров пользователя будет основана на предварительно отфильтрованных наборах данных, которые затем будут использоваться для выражения каждого слова в пользовательском сообщении в виде встроенных слов или в векторной форме (word2vec). Эти наборы данных могут быть предоставлены из Spacy или MITIE, FastText и т.д.;

- контролируемые интенеры (Intent_classifier_tensorflow_embedding). При использовании этого метода пользователю придется создавать данные с нуля, поскольку нет доступных данных обучения.

В связи с специфичной предметной областью и спецификой поиска предварительно обученных наборов данных (большинство предварительно подготовленных данных на английском языке), был создан собственный набор данных обучения. Это гарантирует, что чат-бот будет иметь короткий период обучения и более высокую точность.

Для реализации предлагаемых подходов и их интеграции было разработано чат-бот. Набор вопросов-ответов, собранный из веб-сайтов министерства здравоохранения РФ [13] и Всемирной организации здравоохранения [14]. Общее приложение чат-бота показано на Рисунке 3.

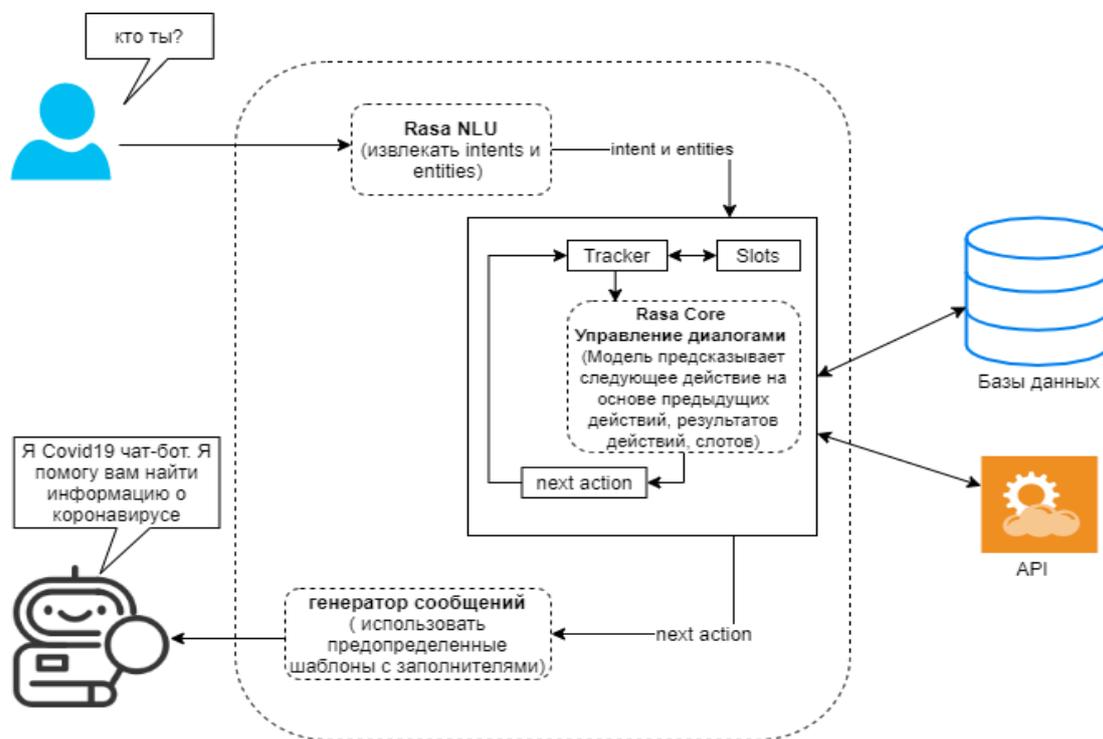


Рисунок 3 - Архитектура чат-бота
 Figure 3 - Chatbot architecture

Когда пользователь формирует запрос, чат-бот обрабатывает его, предсказывает намерение пользователя и дает ответ с наиболее точным намерением. Например, на вопрос «Кому угрожает риск тяжелого течения болезни?» мы получили предсказывание 10-х намерений, но намерение объекта «тяжелого течения болезни» (severeillness) с максимальной уверенностью 0.9758. Поэтому чат-бот будет давать ответ, основанный на намерении объекта «тяжелого течения болезни».

```
{'entities': [],
'intent': {'confidence': 0.9757514595985413, 'name': 'severeillness'},
```

```
'intent_ranking': [{'confidence': 0.9757514595985413, 'name': 'severeillness'},
  {'confidence': 0.07194845378398895, 'name': 'advices'},
  {'confidence': 0.0478818416595459, 'name': 'thank'},
  {'confidence': 0.02501717209815979, 'name': 'greet'},
  {'confidence': 0.0228663831949234, 'name': 'symptoms'},
  {'confidence': 0.008421145379543304, 'name': 'coronavirus'},
  {'confidence': 0.0011407062411308289, 'name': 'give_name'},
  {'confidence': 0.0, 'name': 'COVID-19'},
  {'confidence': 0.0, 'name': 'compliment'},
  {'confidence': 0.0, 'name': 'ask_func'}],
'text': 'Кому угрожает риск тяжелого течения болезни?'
```

Разработанный чат-бот был интегрирован в мессенджер Slack. На Рисунке 4 приведен пример прохождения теста.

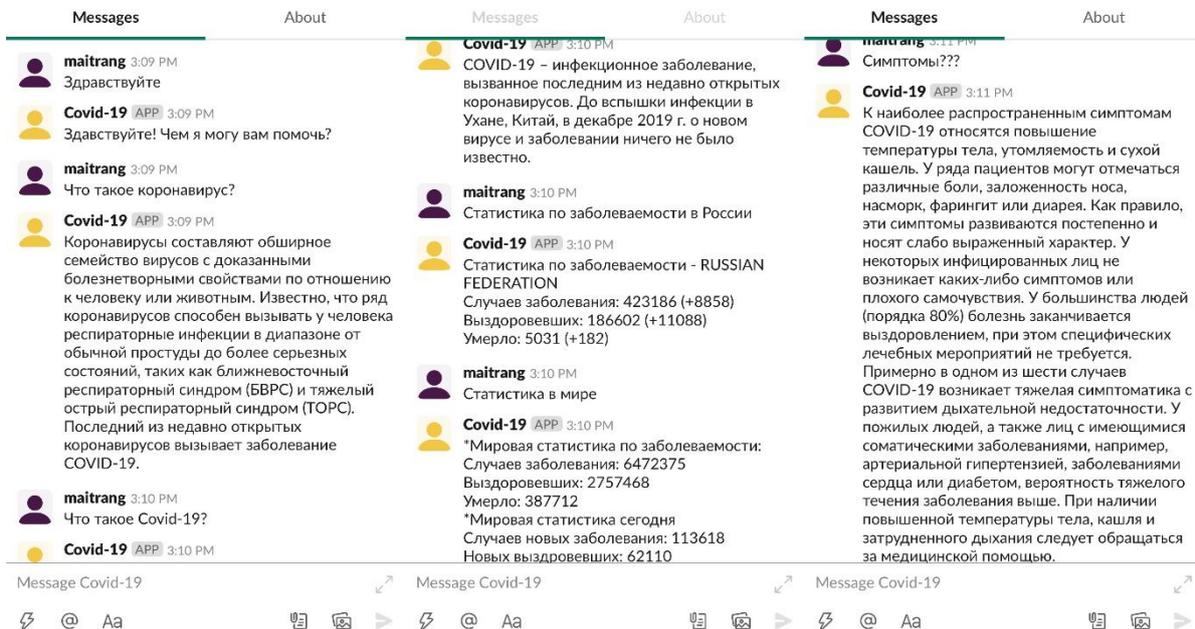


Рисунок 4 - Пример работы чат-бота в мессенджере Slack
Figure 4 - An example of how a chatbot works in the Slack messenger

Тестирование в среде чат-бота было разработано для анализа допустимого ответа по запросу пользователя. Сообщение содержит много информации, включая сущности NLP. Если ключевое слово, используемое в сущности совпадает с правилом тогда появляются сущности NLP. Рассмотрим следующий пример, где сущность NLP в том, что местоположение (location) дается с уверенностью 0.9428 и намерение запроса дается с уверенностью 0.9389.

```
{'entities': [{'confidence': 0.942787306665811,
  'end': 40,
  'entity': 'location',
  'extractor': 'CRFEntityExtractor',
  'start': 34,
```

```
'value': 'России']],
'intent': {'confidence': 0.9389033913612366, 'name': 'request'},
'text': 'статистика количества заражений в России?'
```

Если место, указанное в разговоре, не определено правильно, то оценка доверия стала снижаться. Например, если «в России» написано, как «в Росси», то уверенность сущности снизится с 0.9428 до 0.9173 и уверенность намерения немного меняется с 0.9389 до 0.9381.

```
{'entities': [{'confidence': 0.917320595503475,
               'end': 39,
               'entity': 'location',
               'extractor': 'CRFEntityExtractor',
               'start': 34,
               'value': 'Росси'}]},
'intent': {'confidence': 0.938173770904541, 'name': 'request'},
'text': 'статистика количества заражений в Росси ?'
```

А если «в России» написано, как «в россии», то уверенность сущности снизится с 0.9428 до 0.5145, но уверенность намерения не меняется.

```
{'entities': [{'confidence': 0.5145140534680095,
               'end': 40,
               'entity': 'location',
               'extractor': 'CRFEntityExtractor',
               'start': 34,
               'value': 'россии'}]},
'intent': {'confidence': 0.9389033913612366, 'name': 'request'},
'text': 'статистика количества заражений в россии?'
```

В данном исследовании использованы три различные модели для которых оценивалась результативность.

Базовая модель (B) - это модель, обученная на подготовленных авторами выборок данных. Используемая последовательность обработки данных (pipeline) – конвейер типа supervised embeddings. В этой модели содержат 152 интента.

Базовая модель с синонимами (BS) - модель со списком синонимов, который также является частью обучающей выборки.

Базовая модель с синонимами и шумами (BSS) – модель-расширение модели BS, к которой мы добавили шумы для повышения надежности. Одна случайная строка кириллических символов произвольной длины (от 1 до 10 символов) была добавлена между любыми двумя словами входных предложений. Поскольку не все предложения содержат шум, мы решили добавить случайные строки символов к 10% выборочных данных. Например, такого предложения «Каковы симптомы *флвстто* COVID-19?»

Для оценки результативности мы выбрали модуль оценки Rasa NLU. Модуль оценивает по трем критериям: достоверность (accuracy), точность (precision) и F-мера (в данной статье использовали F1- score) [15].

В таблице 1 представлены результаты оценки для всех трех моделей.

Таблица 1- Оценка трех моделей по критериям: достоверность, точность и F-мера
Table 1- Evaluation of three models by indicators: proportion, accuracy, and F-measure

Модель	Accuracy	Precision	F ₁ - score
B	0.853	0.866	0.834
BS	0.873	0.879	0.857
BSS	0.862	0.884	0.849

Из таблицы видно, что относительно равную результативность моделей. Модель BS с синонимами и модель BSS с синонимами и шумами слов показали результаты лучше по сравнению с базовой моделью.

Заключение

В данной работе рассмотрен подход проектирования чат-бота с использованием фреймворка RASA. С использованием подхода разработан чат-бот. Разработаны три модели: базовая модель (B), базовая модель с синонимами (BS) и базовая модель с синонимами и шумами. Проведена оценка эффективности трех моделей на основании которой можно сделать вывод, что модели BS и BSS результативнее, чем базовая модель.

В настоящее время чат-бот реализован для русского языка. В дальнейшем предполагается многоязычный чат-бот для общения на английском и вьетнамском языке.

ЛИТЕРАТУРА

- ТОП-10 причин, почему чат-боты в соцсетях нужны каждому. Режим доступа: <https://ideadigital.agency/ru/chat-bot/>, (дата обращения 22.03.2020).
- Rasa Stack. Режим доступа: https://legacy-docs.rasa.com/docs/get_started_step1/, (дата обращения 22.12.2019 на англ.).
- Yan M., Castro P., Cheng P., Ishakian V. Building a Chatbot with Serverless Computing. 2016 *Proceedings of the 1st International Workshop on Mashups of Things and APIs (MOTA '16)*. 2016.
- Jyri L., Niko M., Tommi M. Case Study: Building a Serverless Messenger Chatbot. *Current Trends in Web Engineering*. 2018: 75-86.
- Handoyo E., M. Arfan, Y.A.A. Soetrisno, M. Somantri, A. Sofwan, E.W. Sinuraya. Ticketing Chatbot Service using Serverless NLP Technology. 2018 5th International Conference on Information Technology, Computer, and Electrical (ICITACEE). 2018.
- COVID-19. Режим доступа: <https://ru.wikipedia.org/wiki/COVID-19> (дата обращения 22.04.2020).
- Роспотребнадзор назвал высокой вероятностью второй волны COVID-19 осенью. Режим доступа: <https://www.rbc.ru/society/01/05/2020/5eac02069a79474fc26b1648> (дата обращения 01.05.2020).
- Система уведомлений о событиях (Webhooks). Режим доступа: <https://www.unisender.com/ru/support/api/common/sistema-uedomlenij-o-sobytyah-webhooks/> (дата обращения 22.02.2020).
- Rasa NLU: Language Understanding for Chatbots and AI assistants. Режим доступа: <https://rasa.com/docs/rasa/nlu/about/> (дата обращения 22.02.2020 на англ.).
- Markdown. Режим доступа: <https://www.markdownguide.org/> (дата обращения 18.02.2020 на англ.).
- The Rasa Core Dialogue Engine. Режим доступа: <https://rasa.com/docs/rasa/core/about/> (дата обращения 23.02.2020 на англ.).

12. Coronavirus COVID19 API. Режим доступа: <https://documenter.getpostman.com/view/10808728/SzS8rjbc?version=latest> (дата обращения 15.03.2020).
13. Вопросы и ответы коронавируса. Режим доступа : <https://covid19.rosminzdrav.ru/> (дата обращения 25.04.2020).
14. Заболевание, вызванное коронавирусом (COVID-19): Часто задаваемые вопросы. Режим доступа: <https://www.who.int/ru/emergencies/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses> (дата обращения 25.04.2020).
15. A. Zheng. *Evaluating Machine Learning Models*. O'Reilly Media, Inc. 2015.

REFERENCES

1. TOP 10 reasons why everyone needs chatbots in social networks. Available at: <https://ideadigital.agency/ru/chat-bot/> (accessed 22.03.2020).
2. Rasa Stack. Available at: https://legacy-docs.rasa.com/docs/get_started_step1/ (accessed 22.12.2019).
3. Yan M., Castro P., Cheng P., Ishakian V. Building a Chatbot with Serverless Computing. *2016 Proceedings of the 1st International Workshop on Mashups of Things and APIs (MOTA '16)*. 2016.
4. Jyri L., Niko M., Tommi M. Case Study: Building a Serverless Messenger Chatbot. *Current Trends in Web Engineering*. 2018: 75-86
5. Handoyo E., M. Arfan, Y.A.A. Soetrisno, M. Somantri, A. Sofwan, E.W. Sinuraya. Ticketing Chatbot Service using Serverless NLP Technology. 2018 5th International Conference on Information Technology, Computer, and Electrical (ICITACEE). 2018.
6. COVID-19. Available at: <https://ru.wikipedia.org/wiki/COVID-19> (accessed 22.04.2020).
7. Rospotrebnadzor called a high probability of a second wave of COVID-19 in the fall. Available at: <https://www.rbc.ru/society/01/05/2020/5eac02069a79474fc26b1648> (accessed 01.05.2020)
8. System event notifications (Webhooks). Available at: <https://www.unisender.com/ru/support/api/common/sistema-vedomlenij-o-sobytyiah-webhooks/> (accessed 22.02.2020).
9. Rasa NLU: Language Understanding for Chatbots and AI assistants. Available at: <https://rasa.com/docs/rasa/nlu/about/> (accessed 22.02.2020).
10. Markdown. Available at: <https://www.markdownguide.org/> (accessed 18.02.2020).
11. The Rasa Core Dialogue Engine. Available at: <https://rasa.com/docs/rasa/core/about/> (accessed 23.02.2020).
12. Coronavirus COVID19 API. Available at: <https://documenter.getpostman.com/view/10808728/SzS8rjbc?version=latest> (accessed 15.03.2020).
13. Coronavirus Questions and Answers. Available at: <https://covid19.rosminzdrav.ru/> (accessed 25.04.2020).
14. Disease caused by coronavirus (COVID-19): Frequently asked Questions. Available at: <https://www.who.int/ru/emergencies/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses> (accessed 25.04.2020).
15. A. Zheng. *Evaluating Machine Learning Models*. O'Reilly Media, Inc. 2015.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Нгуен Тхи Май Чанг, аспирантка, кафедра «Системы автоматизированного проектирования и поискового конструирования», ФГБОУ ВО «Волгоградский государственный технический университет», Волгоград, Российская Федерация
e-mail: m.trang91@gmail.com

Nguyen Thi Mai Trang, PhD Student, CAD Department, Volgograd State Technical University, Volgograd, Russian Federation

Щербаков Максим Владимирович, д.т.н., профессор, кафедра «Системы автоматизированного проектирования и поискового конструирования», ФГБОУ ВО "Волгоградский государственный технический университет", Волгоград, Российская Федерация.
e-mail: maxim.shcherbakov@gmail.com

Maxim V. Shcherbakov, Dr. Tech. Sci., Professor, CAD Department, Volgograd State Technical University, Volgograd, Russian Federation.