

ВСТРАИВАЕМЫЙ ЧАТ-БОТ В ГОТОВЫЕ КОРПОРАТИВНЫЕ СИСТЕМЫ

Г.С. Чурсин, Н. И. Исмоилов, А. В. Доржиев

Научный руководитель: доцент, к. ф.-м. н. М.Е. Семенов

Национальный исследовательский Томский политехнический университет,

Россия, г. Томск, пр. Ленина, 30, 634050

E-mail: genachursin@yandex.ru

EMBEDDED CHATBOT IN READY CORPORATE SYSTEMS

G. S. Chursin, N. I. Ismoilov, A. V. Dorzhiev

Scientific Supervisor: Assoc. Prof., PhD M. E. Semenov

Tomsk Polytechnic University, Russia, Tomsk, Lenin str., 30, 634050

E-mail: genachursin@yandex.ru

Abstract. In this paper, we consider a corporate chat-bot that can reply to various user commands and uses natural language processing. Moreover, the most common employee's working processes were automated. Our solution can work under any corporate local or global networks. The chat-bot prototype is presented as result.

Введение. В настоящее время идет активное развитие и внедрение интеллектуальных помощников – чат-ботов, которые способны выполнять различные рутинные офисные задачи (поиск информации, первичная обработка данных, консультационная поддержка и оповещение пользователей) получая от пользователей запросы на естественном языке [1]. Распространение и использование чат-ботов привело к парадигме пользовательского взаимодействия messaging-as-an-interface [2]. Наиболее используемой платформой для построения чат-ботов является приложение Telegram, кроме этого отметим также платформы – IBM Watson (IBM), LUIS (Microsoft), Api.ai (Google), Wit.ai (Facebook). Однако, интеграция чат-бота, разработанного с использованием указанных платформ, во внутрикорпоративных системах сталкивается с определенными сложностями, так как в приложении используются скрытые протоколы сбора информации. В связи с чем возникает актуальная задача – разработка компьютерной программы для роботизации бизнес-процессов с использованием только открытого бесплатного программного обеспечения с исходным кодом. Целью данной работы является разработка прототипа чат-бота для интеграции в информационную систему предприятия. Для достижения поставленной цели необходимо последовательно решить следующие задачи:

- а) провести анализ существующих технологий и составить алгоритм работы чат-бота;
- б) разработать библиотеку для обработки естественного языка;
- в) разработать консольный прототип чат-бота;
- г) провести тестирование компонент чат-бота.

Анализ задачи и используемые инструменты. Сотрудники предприятия ежедневно сталкиваются с решением формализованного перечня круга задач. Регулярные потери времени на выполнение рутинных задач, обучение персонала, необходимость поддерживать библиотеки данных, использование перегруженных информацией интерфейсов снижает экономическую эффективность бизнес-процессов. Разработка и внедрение чат-бота позволит автоматизировать решение указанных

задач. Анализ литературы и предметной области позволил сформулировать к чат-боту следующие критерии. Чат-бот должен быть адаптируемым под существующие корпоративные решения, обладать безопасным доступом к информационным ресурсам компании, иметь интуитивно понятный интерфейс пользователя. Создание платформы по обработке команд пользователей может быть реализовано на языке программирования Python или Java, в качестве системы управления базой данных могут быть использованы PostgreSQL, MongoDB или SQLite.

Для разработки чат-бота нами были выбраны следующие программные инструменты.

- Django, Javascript – разработка веб-интерфейса администратора. Административная панель позволяет разграничить права доступа пользователей, настроить периодичность создания архивов, восстановить историю диалогов, планировать и выполнять профилактические работы, отслеживать действия пользователей.

- PostgreSQL – свободная объектно-реляционная система управления базой данных для хранения истории диалогов и запросов.

- Rocket.chat – кроссплатформенное приложение с открытым исходным кодом для корпоративных чатов.

- Natural Language Toolkit (NLTK) библиотека на языке Python для предварительной обработки (токенизация, стемминг, удаление шума и стоп-слов), анализа и выделения ключевых слов пользовательских запросов [3]. На основе этой библиотеки можно проводить лингвистический анализ предложений, то есть выделять ключевые слова и делать предположения о том, какой запрос поступает от пользователя.

- ChatterBot – библиотека генерации ответов на основе машинного обучения [4]. В базу загружаются диалоги на естественном языке. На основе этих диалогов проводится анализ ответов на различные запросы. В последующем, в ходе общения с пользователем, ChatterBot использует образцы этих диалогов для формирования собственных ответов.

Алгоритм готовой системы выглядит следующим образом. Разработанные модули интегрируются в готовые корпоративные решения. В личный кабинет пользователя встраивается дополнительный раздел, где пользователь может получить секретный ключ для авторизации в Rocket.Chat. После регистрации появляется возможность использовать бота. К разработанной программе подключаются корпоративные базы данных, из которых пользователь может получать информацию. Для защищенности доступа к базам данных чат устанавливается на сервер компании.

Результаты. Разработан прототип чат-бота, который через интерфейс командной строки ведет диалог с пользователем на естественном языке. На рис. 1 приведен иллюстративный пример диалога, в ходе которого чат-бот отправил текстовое сообщение. Первая фраза пользователя содержит бытовую лексику с орфографическими ошибками. Для обработки текста были использованы функции `nltk.sent_tokenize()`, `nltk.word_tokenize()`. Для генерации ответов использована библиотека `chatterbot`.

Предварительная обработка текста позволила формализовать запрос пользователя, выделить ключевые слова, затем пользователю последовала команда, в ответ на которую пользователь указал адрес электронной почты. Указанный адрес был автоматически проверен и пользователь получил следующую команду. Текстовое сообщение было успешно отправлено с сохранением синтаксиса и пунктуации.

Тестирование разработанного чат-бота выявили проблемы с поддержкой русского языка: некорректно выделяются ключевые слова (до 80 %), в то время как при использовании английского языка ошибка составляет всего 10-20% в зависимости от сложности запроса. Вероятно, это связано с тем, что nltk не имеет официальной поддержки русского языка, а весь словарь собран сторонними разработчиками.

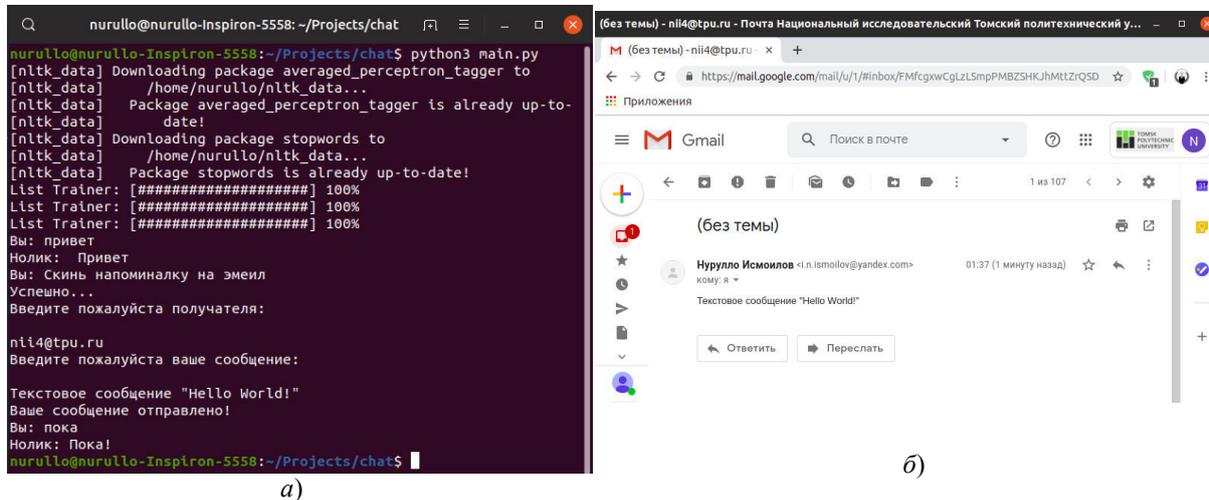


Рис. 1. Диалог с чат-ботом (а), результат отправки текстового сообщения (б)

Заключение. В ходе работы был разработан прототип чат-бота, который способен воспринимать запросы на естественном языке, выделять ключевые слова для выполнения требуемых операций. Выявлено узкое место – обработка запросов на русском языке, потому дополнительно к библиотеке nltk для лемматизации фраз мы использовали библиотеку rymorphy2. В дальнейшем мы планируем подключить к прототипу чат-бота платформу Rocket.Chat, что позволит существенно расширить функционал чат-бота и предоставить пользователю графический интерфейс для обработки как формализованных (справочная информация, юридические запросы), так и неформализованных (восстановление пароля, изменение контактной информации) запросов пользователей, повысить производительность труда и лояльность клиентов. Для генерации релевантных ответов планируется использовать математический аппарат Марковских цепей, что позволит генерировать уникальный текст при сохранении лексики исходного образца [5]. Кроме этого мы планируем реализовать подсистему резервного копирования и защиты от сбоев.

СПИСОК ЛИТЕРАТУРЫ

1. Матвеева Н.Ю., Золотарюк А.В. Технологии создания и применения чат-ботов [Электронный ресурс]. // Научные записки молодых исследователей. – 2018. – №1. – URL: <https://cyberleninka.ru/article/n/tehnologii-sozdaniya-i-primeneniya-chat-botov>
2. Чат-боты и искусственный интеллект в обучении // Журнал EduTech. – 2018. – № 6 (18).
3. NLTK. Официальная документация [Электронный ресурс]: – Режим доступа: <https://github.com/nltk/nltk/wiki/FAQ>. Дата доступа 20.04.2019.
4. Chatterbot. Официальная документация [Электронный ресурс]: – Режим доступа: <https://chatterbot.readthedocs.io/en/stable/index.html>. Дата доступа 20.04.2019.
5. Зорин А.В., Зорин В.А., Пройдакова Е.В. Введение в общие цепи Маркова. Учебно-методическое пособие // Нижегородский госуниверситет – Нижний Новгород, 2013. – 51 с.