

НАПИСАНИЕ ЧАТ-БОТА ПОМОЩНИКА ПОСТУПАЮЩЕМУ В ВЫСШЕЕ УЧЕБНОЕ ЗАВЕДЕНИЕ

WRITING A CHAT BOT ASSISTANT FOR APPLICANTS TO UNIVERSITY

**O. Yudin
I. Yudin**

Summary. This paper proposes a solution to the problem of providing various information to applicants for admission to the university. To solve it, the author proposes a technology for developing a chat bot that functions through an instant messenger. Chat bot uses the search algorithm for synonyms and is able to answer various questions of the applicant related to admission to the university. A prototype of this chatbot for the Telegram message manager was demonstrated. The results of the comparison of the time spent when using a chat bot with the costs when searching for the necessary information on the website of the university, confirming the effectiveness of the use of the bot are given.

Keywords: chat bot, search by synonyms, natural language analysis, messenger, normalization.

Юдин Олег Александрович

Волгоградский Государственный Технический
Университет

Юдин Игорь Александрович

Волгоградский Государственный Технический
Университет
yudin_123@mail.ru

Аннотация. В данной работе предлагается решение проблемы предоставления различной информации абитуриентам при поступлении в ВУЗ. Для её решения автором предлагается технология разработки чат-бота, функционирующего через мессенджер. Чат-бот использует алгоритм поиска по синонимам и способен отвечать на различные вопросы абитуриента, связанные с поступлением в ВУЗ. В работе продемонстрирован прототип данного чат-бота для мессенджера Telegram. Приводятся результаты сравнения затрачиваемого времени при использовании чат-бота с затратами при поиске необходимой информации на сайте ВУЗа, подтверждающие эффективность использования бота.

Ключевые слова: чат-бот, поиск по синонимам, анализ естественного языка, мессенджер, нормализация.

В настоящее время рынок чат-ботов стремительно растет. Почти у каждого крупного предприятия есть собственный «умный помощник» в используемых приложениях для мгновенного обмена сообщениями, здесь и далее именуемых мессенджерами [1]. Чат-боты предназначены для ответов на вопросы пользователей: клиентов, посетителей сайтов, партнеров и т.п. Мессенджерами для мобильных устройств пользуются почти 60% обладателей смартфонов в России, что составляет 50–60 млн. человек, по оценкам J'son & Partners Consulting. 10% пользователей мессенджеров признают, что регулярно общаются с ботами [1].

Боты привлекательны тем, что способны заменить людей при выполнении различных рутинных задач, тем самым экономя средства предприятия. Бота можно создать один раз и использовать бессрочно, периодически обновляя его базу данных. Потенциально он может заменить бизнес-ассистента или секретаря с зарплатой 50 тыс. руб. в месяц [1].

Современные учебные заведения стремятся повысить уровень качества образования и его интерактивность, а также поддерживать постоянный поток абитуриентов. Как следствие большинство ВУЗов имеет свои веб-сайты с информацией, необходимой студентам, абитуриентам и их родителям. Часто информация

на данных веб-сайтах не является актуальной, а сам сайт может иметь запутанную структуру и быть сложным для восприятия ввиду зачастую малых бюджетов на его поддержку и улучшение оформления. Это значительно повышает временные затраты и усложняет процесс поиска информации на сайте для принятия решений. Одним из возможных способов решения данной проблемы может быть создание чат-бота ВУЗа, функционирующего через один из мессенджеров. Используя чат-бот, пользователь может в короткие сроки получить необходимую ему информацию вместо того, чтобы искать ее на сайте [6].

На сегодняшний день в обработку естественного языка входит распознавание речи, классификация, экстракция знаний из текстов и другие действия, направленные на понимание текстов с целью наполнения баз знаний, формирования ответов на вопросы и ведения диалога. Первые подходы к глубокой обработке текстов обычно сводились к разбору языка с применением грамматик. Также это делалось путем перевода из дерева разбора в некоторое логическое представление знаний с помощью свода правил и специально заготовленного лексикона. После этого логическое представление можно было добавить в базу знаний и выполнять над ней разного рода операции, отвечать на вопросы, проверять утверждения и т.п. Однако при попытках практи-

ческого применения этого подхода возникали сложности, связанные в том числе с необходимостью учитывать общепринятые (то есть элементарные, базовые) знания о мире [4].

Первые чат-боты появились еще в 1990-е годы. Они использовались в IRC (Internet Relay Chat, протоколах прикладного уровня) для обмена сообщениями. Сейчас чат-боты распознают человеческую речь и способны поддержать диалог [1].

Боты бывают двух типов. Простые древовидные чат-боты выполняют только определенные команды. В отличие от старых разработок они могут получать информацию из сторонних ресурсов, самостоятельно искать ответы в сети Интернет. У таких чат-ботов может быть графический интерфейс. Древовидные чат-боты реагируют на нажатие кнопок на экране, предлагают списки для выбора вариантов, отображают товары [1]. Чат-боты второго, нового типа, «понимают» собеседника, постоянно обучаются и могут отвечать не по шаблону. В разработке таких чат-ботов используются специальные технологии, предназначенные для распознавания текста и голоса [1].

Рассмотрим несколько существующих чат-ботов российских ВУЗов, предоставляющих студентам интересующую их информацию. Одним из них является BFUBot — бот Балтийского Федерального Университета имени Иммануила Канта, функционирующий через мессенджер Telegram [2]. На текущий момент чат-бот может выполнить следующие функции:

- ◆ отправить имя пользователя и пароль для доступа к информационным системам университета, студенческой электронной почте;
- ◆ показать пароль беспроводной сети WiFi;
- ◆ получить предложения от студентов по развитию информационных систем и сервисов БФУ им. И. Канта (механизм адресной обратной связи);
- ◆ получить номер карты и номер ISIC;
- ◆ заказать справки студентами;
- ◆ подать заявки на обслуживание информационной инфраструктуры (механизм технической поддержки обучаемых);
- ◆ подписаться на рассылки и уведомления о деятельности университета (механизм подключения к уведомлениям о мероприятиях, спортивных событиях, новостях университета);
- ◆ получить статистику по прохождению занятий по физической культуре.

Чат-бот для идентификации обучаемого использует номер телефона, который был предоставлен в приемную комиссию при поступлении. Из его недостатков

можно выделить то, что он не предлагает помощи поступающим в него абитуриентам в вопросах, связанных с зачислением. Также чат-бот не реагирует на текстовые сообщения пользователя, общаясь с ним только посредством команд.

Далее был рассмотрен бот `printschedule_test_bot`, используемый в ОГУ имени И. С. Тургенева, также функционирующий через Telegram [3]. Данный чат-бот разрабатывался для предоставления обучающимся информации по их расписанию. Ответ чат-бота содержит следующую информацию:

- ◆ порядковый номер занятия;
- ◆ время занятия;
- ◆ название предмета;
- ◆ тип занятия;
- ◆ учебный корпус;
- ◆ номер аудитории;
- ◆ ФИО преподавателя.

Для предоставления данной информации чат-бот запрашивает от пользователя номер курса, номер группы и день недели, на который необходимо предоставить расписание.

Данный чат-бот как и бот Балтийского университета не предлагает помощи поступающим абитуриентам в вопросах, связанных с зачислением. Также чат-бот не реагирует на текстовые сообщения пользователя, общаясь с ним только посредством команд.

Таким образом, решение для помощи абитуриентам при зачислении отсутствует. Актуальной и востребованной задачей является разработка чат-бота, реализующего данный функционал, осуществляющего диалог с пользователем на естественном языке. В данной работе представлен чат бот, который решает следующие задачи:

- ◆ предоставление актуальной информации для абитуриентов с возможностью задания интересующих вопросов про поступление в ВолгГТУ;
- ◆ снижение временных затрат на поиск информации за счет ее предоставления пользователю вместо поиска на сайте вручную.

В чат-боте, предлагаемом к рассмотрению в данной работе, были использованы следующие алгоритмы:

- ◆ алгоритм обработки сообщения пользователя;
- ◆ алгоритм обработки вопросов из базы данных;
- ◆ алгоритм поиска наиболее подходящего ответа;
- ◆ алгоритм подготовки используемых документов;
- ◆ алгоритм работы с командами.

Для реализации алгоритма обработки сообщения пользователя были разработаны следующие функции:

- ◆ поиск и устранение ошибок в строке;
- ◆ приведение каждого слова к нормальной форме;
- ◆ удаление незначимых частей речи.

Функция поиска и устранения ошибок в строке получает на вход исходное сообщение пользователя и удаляет из него все слова и не кириллические символы. Затем осуществляется проверка каждого слова сообщения на наличие грамматических ошибок.

Для корректного поиска ответа необходимо учитывать, что слова в русском языке могут менять свою форму. Поэтому необходимо каждое слово привести к его нормальной форме [5]: для существительного — к единственному числу и именительному падежу, для глагола — к инфинитиву, для прилагательного — к единственному числу, именительному падежу.

На вход функции приведения каждого слова к нормальной форме передается сообщение пользователя. Она разбивает его на токены (отдельные слова), вызывает функцию приведения слова к нормальной форме и возвращает список слов в нормальной форме.

Для повышения точности поиска ответа необходимо избавиться от таких частей речи как союзы, предлоги, междометия, местоимения, частицы. Функция удаления не значимых частей речи получает на вход список слов, удаляет из списка все не значимые части речи и возвращает обработанный список.

Для корректного поиска по совпадающим словам необходимо также обработать вопросы из базы данных. Для реализации данного алгоритма была разработана функция приведения слов из вопросов базы данных к нормальной форме. Данная функция разбивает вопросы из базы данных на отдельные слова и вызывает функцию приведения слов к нормальной форме. На выходе формируется вложенный список слов.

Для реализации данного алгоритма были разработаны следующие функции:

- ◆ поиск синонимов слова;
- ◆ поиск синонимов слова из сообщения пользователя в вопросах из базы данных;
- ◆ поиск релевантного ответа.

Для повышения точности поиска релевантного ответа необходимо, чтобы учитывались синонимы каждого слова из сообщения пользователя. Функция поиска синонимов слова получает на вход слово, для которого необходимо найти синоним. Для данного слова из словаря синонимов ищутся все его синонимы, которые функция возвращает в виде списка.

Функция поиска синонимов слова из сообщения пользователя в вопросах из базы данных получает на вход слово из сообщения пользователя и обработанный вопрос из базы данных в виде списка нормализованных слов. Функция внутри себя вызывает функцию поиска синонимов для слова из сообщения пользователя, которая выдает список синонимов для переданного слова. Функция перебирает каждое слово из списка синонимов и определяет есть ли оно в обработанном списке пользователя. Если оно есть, то это слово возвращается и заносится в список общих слов для сообщения пользователя и вопросов из базы данных.

Функция поиска релевантного ответа получает на вход сообщение пользователя и последовательно вызывает функции устранения ошибок в строке, приведения каждого слова к нормальному виду, удаления не значимых частей речи. Далее функция вызывает функцию поиска синонимов слова из сообщения пользователя в вопросах из базы данных для каждого слова из обработанного сообщения пользователя. Формируется список найденных слов для каждого вопроса из базы данных. Определяем, для какого вопроса из базы данных было найдено больше всего общих слов, и считаем, что данный вопрос является наиболее похожим на заданный, выдаем ответ из базы данных, соответствующий данному вопросу. Если ни в одном вопросе из базы данных не было найдено общих слов с обработанным сообщением пользователя, функция возвратит строку, в которой будет написано, что ответ непонятен. Если подходят несколько вопросов из базы данных, то функция вернет список подходящих вопросов и соответствующие им команды.

Для реализации загрузки и обработки используемых документов была реализована функция загрузки словаря синонимов, а также функция загрузки и обработки таблицы вопросов и ответов.

Для реагирования на команды был реализован алгоритм работы с командами. При первом обращении к чат-боту можно только отправить команду start, начинающую диалог и выводящую приветственное сообщение. Таким образом общение с ботом начинается start-командой. При получении данной команды бот выдает приветственное сообщение. В приветственном сообщении бот сообщает, что для ознакомления со списком все доступных команд необходимо воспользоваться командой help. При вводе пользователем данной команды, бот выдает список всех доступных команд. При вводе пользователем любой другой доступной команды, бот выдает ответ из базы данных, соответствующий введенной команде.

Для реализации предлагаемого чат-бота был выбран мессенджер Telegram [3] — бесплатный кроссплатформенный

Таблица 1. Пример словаря синонимов

инспирация	Инспирирование, наущение, подстрекание, заводка, вдох, выдох, подстрекательство, подначка, внушение, подговор
инспирировавший	Подстрекнувший, внушивший, наущавший, внушавший, инициировавший, влиявший, вызвавший, подстрекавший, мутивший, волновавший, вызывавший
инспирировавшийся	Инициировавшийся, вызвавшийся, мутившийся, вызывавшийся, подстрекавшийся
инспирировал	вдохновлял
инспирировала	вдохновляла

Таблица 2. Пример базы данных вопросов и ответов

«Где посмотреть рейтинг абитуриента?»	«Рейтинг абитуриентов можно увидеть, перейдя по ссылке http://welcome.vstu.ru/acceptance/rejting-abiturientov/ »
«Телефон приемной комиссии»	«(8442) 23–22–92»
«Адрес университета»	«Россия, 400005, Волгоград, пр. им. Ленина, 28»
«Контакты приемной комиссии»	«Россия, 400005, Волгоград, пр. им. Ленина, 28, ауд. 211. Режим работы: Пн-Пт: 10:00–17:00 Перерыв: с 12.30–13.00»

Таблица 3. Пример команд и ответов

«Рейтинг университета?»	«66 место среди ВУЗов России»	/u_raiting
«Сколько ехать от общежития до университета по времени?»	«Время пути до университета приблизительно 15 минут»	/time
«Как получать стипендию?»	первый семестр стипендия выплачивается всем первокурсникам, после вы должны сдать сессию в основные сроки без единой тройки на экзаменах, и набрать средний балл не менее 76 с учетом экзаменов	/get_scholarship
«Какой рейтинг у университета?»	66 место среди ВУЗов России	/u_raiting

менный мессенджер для смартфонов и других устройств, позволяющий обмениваться текстовыми сообщениями и медиафайлами различных форматов.

TelegramBotAPI — библиотека, созданная для разработки ботов для Telegram. Предоставляет возможности создания ботов, отправки сообщений, создания декораторов, получения информации о сообщении пользователя [7]. Telegram предоставляет возможность для пользователей использования команд. Команда — это строка, начинающаяся с символа косой черты «/» и состоящая из символов латинского алфавита и цифр и подчеркиваний. Команда не может содержать больше 32 символов.

Сообщение пользователя не может содержать в себе больше 50 символов. На вход чат-бот принимает только текстовое сообщение или команду. В случае некорректно введенной информации или невозможности чат-ботом понять введенное, выдается сообщение об этом.

Для решения задачи нормализации была использована библиотека rutmorphy2. Для определения кор-

ректности слов в введенном пользователем сообщении используется библиотека ruaspeller. В случае если слово корректно, его оставляем без изменения, если некорректно, то находим наиболее вероятную правильную форму слова и заменяем им начальное слово.

Словарь синонимов хранится на сервере в формате TXT. Первое слово каждой строки выступает в роли ключа, все остальные слова, включая первое, являются значениями для ключа. Пример словаря синонимов показан в таблице 1.

База данных вопросов и ответов хранится на сервере в формате CSV. Для загрузки базы данных используется функция read_csv библиотеки pandas. Пример базы данных вопросов и ответов показан в таблице 2. Пример команд и ответов на них показан в таблице 3.

Разработанный чат-бот был интегрирован в мессенджер Telegram. Для начала работы с чат-ботом необходимо в поиске Telegram набрать имя бота — VolBot. При вводе команды start появляется приветственное сообщение с пояснением работы чат-бота. В сообщении так-

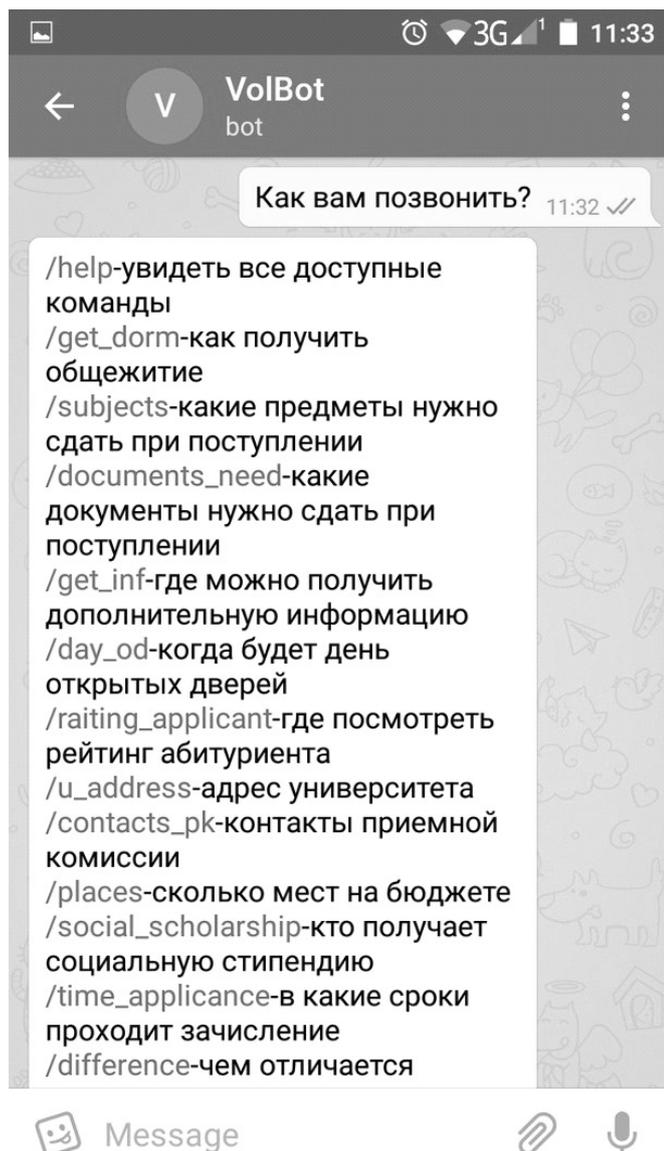


Рис. 1. Ответ чат-бота вопрос пользователя «Как вам позвонить?» (экран 1)

же показывается команда help, которая при вводе должна выводить список всех доступных команд.

При вводе вопроса или известной команды бот выдает ответ на вопрос, а также список всех доступных пользователю команд, чтобы упростить поиск необходимой информации. К примеру, при вводе вопроса «Как вам позвонить?» бот должен вывести номер приемной комиссии, а также список команд, которые могут заинтересовать пользователя. На рисунках 1 и 2 представлен ответ чат-бота на вопрос пользователя «Как вам позвонить?».

На рисунках 1 и 2 видно, что при вводе пользователем сообщения «Как вам позвонить?» бот выдает номер

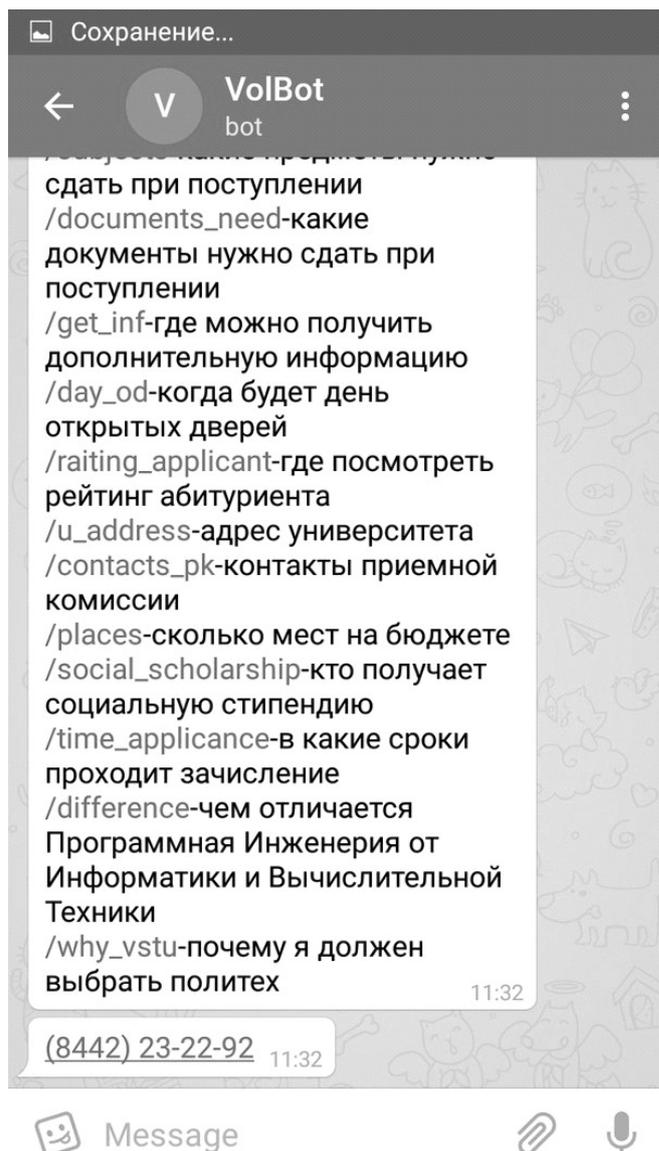


Рис. 2. Ответ чат-бота на вопрос пользователя «Как вам позвонить?» (экран 2)

телефона приемной комиссии, а также список команд, которые могут быть интересны пользователю. При этом важно заметить, что в базе данных не содержится вопрос сформулированный именно так, поиск в этом случае осуществляется по синонимам.

Разработанный чат-бот для Telegram направлен на помощь абитуриентам в поиске необходимой информации при поступлении в ВолгГТУ. Бот способен как отвечать на вопрос пользователя, сформулированный на естественном языке, так и работать с командами. Для проверки изменения затрат времени на поиск необходимой информации был проведен ряд тестирований. Группа из 10 студентов получила задание найти ответы

на сформулированные вопросы о поступлении в ВолгГТУ сначала на сайте ВУЗа, а затем работая с чат-ботом. В результате работа с чат-ботом сократила затраты времени в среднем на 30%.

Таким образом в данной работе была рассмотрена проблема поиска необходимой абитуриенту информации с помощью чат-бота, предложен алгоритм

чат-бота, реализующего поиск релевантного ответа по синонимам, разработан прототип чат-бота, функционирующего через Telegram и способного отвечать на вопросы, возникающие при поступлении в ВолгГТУ и проведено тестирование, подтверждающее, что работа с чат-ботом является более быстрым способом поиска необходимой абитуриентам информации чем поиск на сайте.

ЛИТЕРАТУРА

1. Гайсина И. Бум ботов: как роботы вытесняют живых сотрудников // РБК. 2016. URL: https://www.rbc.ru/own_business/11/10/2016/57fd4b09a79470fc8973 (дата обращения 22.03.2019)
2. БФУ им. И. Канта теперь в Telegram // Официальный сайт Балтийского Федерального Университета имени Иммануила Канта. 2016. URL: <https://www.kantiana.ru/news/142/199690/> (дата обращения 22.03.2019).
3. Козлов А. А., Батищев А. В. Телеграм-бот как простой и удобный способ получения информации // Территория науки. 2017. № 5. URL: <https://cyberleninka.ru/article/n/telegram-bot-kak-prostoy-i-udobnyy-sposob-polucheniya-informatsii> (дата обращения: 10.03.2019).
4. Велихов П. Машинное обучение для понимания естественного языка // Открытые системы. 2016. URL: <https://www.osp.ru/os/2016/01/13048649/> (дата обращения 29.03.2019).
5. Козиев И. Нормализация и лемматизация текста с использованием тезауруса // Инструменты для NLP разработчика: лексика, морфология, синтаксис русского языка. 2019. URL: http://www.solarix.ru/for_developers/docs/text-normalization.shtml (дата обращения 25.03.2019).
6. Чат-бот или веб-сайт: пять основных преимуществ // Medium. URL: <https://medium.com/@hotbot/чат-бот-или-веб-сайт-пять-основных-преимуществ-1e61ea1e7123/> (дата обращения 23.03.2019).
7. Работа с библиотекой PyTelegramBotAPI // Groosha. URL: https://groosha.gitbooks.io/telegram-bot-lessons/content/pytelegrambotapi_basics.html (дата обращения 25.03.2019).

© Юдин Олег Александрович, Юдин Игорь Александрович (yudin_123@mail.ru).
Журнал «Современная наука: актуальные проблемы теории и практики»



Мамаев курган, г. Волгоград