

6. Положение о порядке проведения государственной экспертизы средств технической и криптографической защиты информации [Электронный ресурс] : утв. приказом Оперативно-аналитического центра при Президенте Респ. Беларусь от 26 авг. 2013 г. № 60 : в ред. от // Эталон-Беларусь / Нац. цент правовой информ. Респ. Беларусь. – Минск, 2020.

7. Положение о порядке аттестации систем защиты информации информационных систем, предназначенных для обработки информации, распространение и (или) предоставление которой ограничено, не отнесенной к государственным секретам [Электронный ресурс] : утв. приказом Оперативно-аналитического центра при Президенте Респ. Беларусь, 30 авг. 2013 г. № 62 : в ред. от 11 октября 2017 г. № 64 // Эталон-Беларусь / Нац. цент правовой информ. Респ. Беларусь. – Минск, 2020.

УДК 004.89

Мычко Н. А.

Научный руководитель: к.т.н., доцент Крапивин Ю. Б.

МЕТОДЫ АВТОМАТИЗАЦИИ ОБРАБОТКИ ТЕКСТОВЫХ ОБРАЩЕНИЙ В СЛУЖБУ ТЕХНИЧЕСКОЙ ПОДДЕРЖКИ

Увеличение использования технологий в современности приводит к росту требований к обеспечению службы технической поддержки (далее – СТП). Техническая поддержка часто подразделяется на уровни с целью улучшения обслуживания организации или базы клиентов. Количество уровней определяется потребностями и желаниями бизнеса или же ставится в зависимость от возможностей эффективно помочь клиентам или пользователям [1].

В свою очередь высокая скорость роста количества пользователей в той или иной сфере вынуждает совершенствовать способы обработки обращений, реализовывать частичную или полную автоматизацию этих процессов. Легко заметить, что основным и наиболее востребованным способом представления информации является текст на естественном языке. Именно поэтому далее речь пойдёт об обработке подобных обращений [2].

Обобщенное представление процесса автоматического анализа текста в контексте практически любой задачи, связанной с автоматизацией обработки текстовых сообщений, схематично может быть представлено на различных уровнях, например так, как это сделано на рисунке 1.

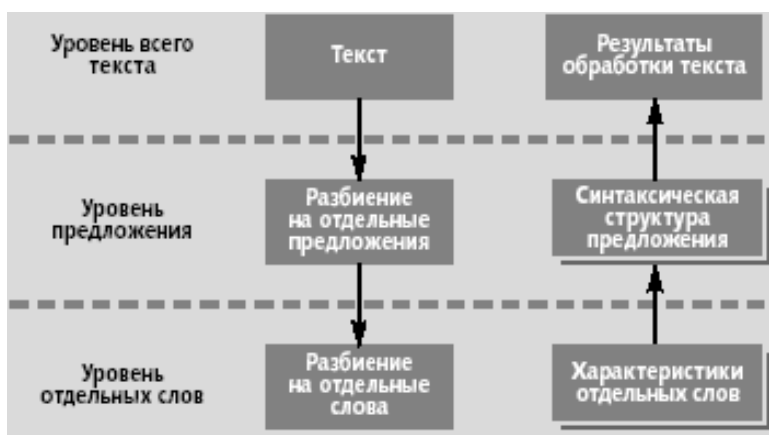


Рисунок 1 – Уровни автоматического анализа текста

В современных информационных системах, как правило, этот процесс реализуется с помощью специального отдельного модуля – лингвистического процессора, выполняющего следующие основные этапы анализа:

1. Предварительное форматирование текста (преформатирование).
2. Лексический анализ.
3. Лексико-грамматический анализ.
4. Синтаксический анализ.
5. Семантический анализ [3].

Существует несколько методов обработки текстовых обращений в контексте их реализации для службы технической поддержки:

1. Ручная обработка от специалиста СТП. Данный метод имеет место в небольших компаниях, т. к. при увеличении количества пользователей нужно либо увеличивать штат работников, что в свою очередь дорого, либо увеличивать время ожидания ответа.

2. Частичная автоматизация А-20: система, автоматизированная на 20%. Ключевой особенностью такой системы является так называемый раздел FAQ (с англ. frequently asked questions, или часто задаваемые вопросы). Его наличие обеспечивает достаточно быстрый ответ в случае, когда проблема простая или действительно часто встречается, ведь именно ответы на такие обращения содержит FAQ.

3. Частичная автоматизация А-40: система, автоматизированная на 40%, кроме раздела FAQ может предложить пользователю просмотреть раздел «форум», где возможно существуют схожие обращения от других клиентов. Подобные системы являются очень полезными в условиях, когда времени на ожидание специалиста мало или нет вовсе. Однако на форумах, где есть возможность ответа не только специалистов, но и обычных пользователей, появится риск появления не совсем корректных решений.

4. Частичная автоматизация А-70: система, автоматизированная на 70% может работать с использованием чат-бота, обученного на предыдущих запросах и ответах для генерации возможных вариантов решения, имеет базовый набор команд, однако всё равно требует присутствия специалиста СТП, т. к. бот не способен решить новые проблемы или проблемы, состоящие из нескольких неполадок. Недостатком такой системы является неспособность предложить решение, выходящее за рамки тех проблем, которые возникали ранее. Преимуществом является их стоимость, возможность оперативно уведомить пользователя о возможном решении проблемы.

5. Частичная автоматизация А-90: системы, автоматизированные на 80-90%, используют внешние сервисы и программы.

Достаточно популярными в настоящее время являются решения на базе машинного обучения и API: Microsoft Language Understanding – это облачная служба, которая позволяет проводить анализ текста на естественном языке, определяя намерения, настроение, извлекая информацию о заданных объектах и пр. Работает оно по принципу разбиения запроса на составные части и предоставления данных в виде ответов JSON [4]. Структура чат-бот приложения, спроектированного на базе облачной службы Microsoft Language Understanding, представлена на рисунке 2.

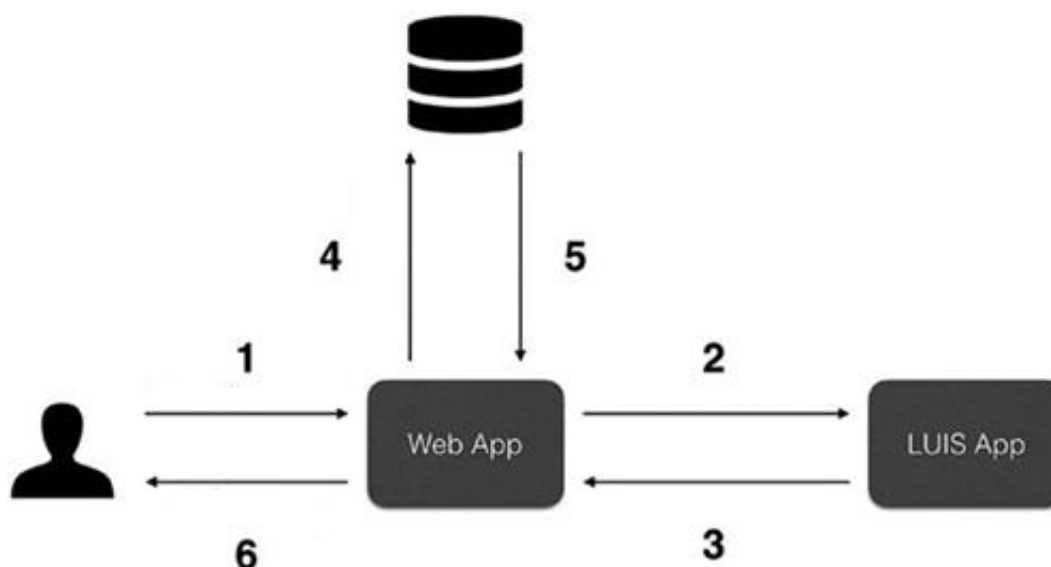


Рисунок 2 – Структура чат-бот приложения, спроектированного на базе облачной службы Microsoft Language Understanding

В соответствии с представленной схемой пользователь формирует запрос в виде текстового сообщения (1), которое поступает в блок «Web App» – веб-приложение, обеспечивающее интерфейс с пользователем и реализующее логику обработки запросов такого плана, далее, (2) передается в блок «LUIS App», где сосредоточена логика его анализа на базе машинного обучения. Результат в виде JSON-файла, содержащий всю необходимую для реализации логики работы приложения информацию, например, тип сообщения, его характер, интересующие пользователя сведения, автоматически извлеченные из сообщения, далее (3) снова поступают в блок «Web App», где выступают в качестве входных данных для поискового запроса к базе данных (4), и сформированный результат (5), содержащий развернутый ответ, также в виде текстового сообщения возвращается пользователю (6).

Google Natural Language API - этот сервис позволяет работать с функциями понимания естественного языка, включая анализ настроения, классификацию контента и синтаксический анализ. Выборочно извлекает ключевые данные для идентификации важных моментов в тексте, многие из которых не отображаются в стандартных языковых моделях [5].

Программный комплекс Naumen обладает такими возможностями, как автоматическая классификация обращения с учетом тематики, приоритета, порядка рассмотрения и других заданных характеристик, возможность назначения ответственных сотрудников с учетом различных параметров [6].

В системах, автоматизированных на 90%, для обработки можно использовать такие библиотеки машинного обучения с открытым кодом, как TensorFlow [7] и Keras [8]. Благодаря этим библиотекам можно в достаточно краткие сроки создать собственную систему. Системы, разработанные по индивидуальным заказам с применением чат-ботов и подходов на базе машинного обучения с помощью глубоких нейронных сетей, имеют ряд преимуществ, как правило: быстрый отклик; малый процент неудачных ответов; экономия людских и финансовых ресурсов при достаточно высоком качестве услуг. В качестве недостатков отмечаются необходимость в больших наборах обучающих данных, сложность разработки.

Существующие методы автоматизации обработки текстовых обращений в службу технической поддержки требуют достаточно наукоемких решений и могут обеспечивать высокое качество информационных услуг при относительно небольших финансовых и временных затратах в случае использования доступных наработок в этой сфере.

Список цитированных источников

1. Википедия [Электронный ресурс] : Техническая поддержка - Режим доступа: <https://ru.wikipedia.org/wiki/> - Дата доступа: 23.05.2020.
2. CyberLeninka [Электронный ресурс] : Автоматизация обработки текстов естественного языка - Режим доступа: <https://cyberleninka.ru/> - Дата доступа: 23.05.2020.
3. Открытые системы [Электронный ресурс] / Обработка текстов на естественном языке – Режим доступа: <https://www.osp.ru/os/> - Дата доступа: 04.05.2020.
4. Language Understanding [Электронный ресурс] : Documentation - Режим доступа: <https://www.luis.ai/> - Дата доступа: 23.05.2020.
5. Google cloud [Электронный ресурс] : Natural Language - Режим доступа: <https://cloud.google.com/> - Дата доступа: 23.05.2020.
6. Naumen [Электронный ресурс] : Интеллектуальная обработка обращений - Режим доступа: <https://www.naumen.ru/> - Дата доступа: 23.05.2020.
7. TensorFlow [Электронный ресурс] : TensorFlow - Режим доступа: <https://www.tensorflow.org/> - Дата доступа: 23.05.2020.
8. Keras [Электронный ресурс] : Keras Documentation - Режим доступа: <https://keras.io/> - Дата доступа: 23.05.2020.

УДК 004.94

Онысько А. А.

Научный руководитель: к.т.н., доцент Хведчук В. И.

СИСТЕМА ОЦЕНКИ КАЧЕСТВА ПЕРЕВОДА С ВОЗМОЖНОСТЬЮ ГОЛОСОВОГО ВВОДА И ВЫВОДА

Введение. Перевод играет огромную роль в современном мире, особенно после того, как произошел так называемый «информационный взрыв». Резко увеличился объем информации, которой обмениваются люди и народы. Возросли международные контакты, на карте мира появились новые государства, возникли многочисленные международные организации, всемирные движения, региональные союзы государств. Научно-техническая революция вызвала огромную потребность в обмене научной информацией между разными странами.

Цель работы – создание системы, позволяющей выполнить перевод на многие современные языки с возможностью голосового ввода и вывода, также будет осуществляться лексический и синтаксический контроль. Актуальность данной системы в современном мире очень высока, многие сферы в жизни требуют от человека знания как минимум одного иностранного языка, а если говорить про сферу ИТ, то без иностранного языка здесь вообще никак. Не каждый человек идеально знает не свой родной язык, из-за этого возникают проблемы с коммуникацией между заказчиком, или, например, коллегой, что приводит к недопониманию, неправильной трактовке требования и т. д. Разрабатываемая система поможет решить эти проблемы.

1. Обзор аналогов

SYSTRAnet

Так называемый сервис-старожил, который более 40 лет предлагает свои