

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ**»

Инженерная школа ядерных технологий
Направление подготовки: Прикладная математика и информатика
Отделение экспериментальной физики

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Встраиваемый чат-бот в готовые корпоративные системы

УДК 519.237:004.7.032.26:159.242

Студент

Группа	ФИО	Подпись	Дата
ОВМ71	Чурсин Геннадий Сергеевич		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Семенов М.Е.	к.ф-м.н.		

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Меньшикова Е.В.	к.ф.н		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Исаева Е.С.			

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Профессор	Трифонов А.Ю.	д.ф-м.н.		

Томск – 2019 г.

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа ядерных технологий

Направление подготовки 01.04.02 «Прикладная математика и информатика»

Отделение экспериментальной физики

УТВЕРЖДАЮ:

Руководитель ООП

(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации

Студенту:

Группа	ФИО
0ВМ71	Чурсину Геннадию Сергеевичу

Тема работы:

Встраиваемый чат-бот в готовые корпоративные системы	
Утверждена приказом директора (дата, номер)	

Срок сдачи студентом выполненной работы:	
--	--

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.);</i>	Объект исследования – чат-бот; Предмет исследования - естественный язык. Библиотека автоматизирует выполнение рутинных офисных задач
Перечень подлежащих исследованию, проектированию и разработке вопросов <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i>	<ol style="list-style-type: none">1. Изучить используемые алгоритмы2. Написать библиотеку для обработки естественного языка в рамках задачи3. Разработать модуль обработки опечаток4. Разработать модули, решающие офисные задачи (поиск информации, заполнение документов и пр.)5. Провести тестирование
Перечень графического материала <i>(с точным указанием обязательных чертежей)</i>	Результаты ответов чат-бота
Консультанты по разделам выпускной квалификационной работы <i>(если необходимо, с указанием разделов)</i>	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Меньшикова Е.В.
Социальная ответственность	Исаева Е.С.

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	
---	--

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Семенов М.Е.	к.ф.-м.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0ВМ71	Чурсин Геннадий Сергеевич		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
0BM71	Чурсину Геннадия Сергеевичу

Школа	Инженерная школа ядерных технологий	Отделение (НОЦ)	Отделение экспериментальной физики
Уровень образования	Магистратура	Направление/специальность	01.04.02 Прикладная математика и информатика

Тема ВКР:

Встраиваемый чат-бот в готовые корпоративные системы	
Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Произведен анализ рабочего места человека, который будет работать с полученной в ВКР разработкой. Данная разработка окажет огромную помощь в выполнении рутинных задач сотрудниками. В работе будет описан перечень используемых характеристик, а также унификация терминологии данной сферы деятельности. Сократится время на поиски нужной информации.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Правовые и организационные вопросы обеспечения безопасности: – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны.	– ТК РФ от 30.12.2001 N 197-ФЗ (ред. от 01.04.2019) – ГОСТ 12.2.032-78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования».
2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов	В качестве вредных факторов выделены: повышенный уровень электромагнитных излучений, недостаточная освещенность рабочей зоны, нервно-психические

2.2. Обоснование мероприятий по снижению воздействия	перегрузки, отклонения микроклимата.
3. Экологическая безопасность:	Рассматриваются воздействия на окружающую среду при утилизации средств измерений и электрических приборов.
4. Безопасность в чрезвычайных ситуациях:	Существует вероятность возникновения техногенных ЧС, особенно пожаров.

Дата выдачи задания для раздела по линейному графику	
--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель ООД	Исаева Елизавета Сергеевна			

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0ВМ71	Чурсин Геннадий Сергеевич		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа	ФИО
0BM71	Чурсину Геннадию Сергеевичу

Школа	ИЯШТ	Отделение школы (НОЦ)	Экспериментальной физики
Уровень образования	Магистр	Направление/специальность	01.04.02 Прикладная математика и информатика

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): <i>материально-технических, энергетических, финансовых, информационных и человеческих</i>	З/П – 77593,56 Всего бюджет - 161152,45
2. Нормы и нормативы расходования ресурсов	1. Ставка 1 кВт на электроэнергию – 5,8 руб.
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	1. Коэффициент отчислений на уплату во внебюджетные фонды - 27,1 2. Район. Коэф – 30% 3. Накл расх -30%

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого и инновационного потенциала НТИ	Технико-экономическое обоснование НИР
2. Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок	Разработка диаграммы Ганта, разработка этапов проведения работ, расчет бюджета
3. Определение ресурсной, финансовой, экономической эффективности	Оценка научно-технического уровня НИР

Перечень графического материала (с точным указанием обязательных чертежей):

Дата выдачи задания для раздела по линейному графику	
--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент	Меньшикова Екатерина	к.ф.н		

	Валентиновна			
--	--------------	--	--	--

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0ВМ71	Чурсин Геннадий Сергеевич		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ООП

Код результата	Профессиональные компетенции
	<i>1) В области научно-исследовательской деятельности:</i>
(ПК-1)	Способность проводить научные исследования и получать новые научные и прикладные результаты самостоятельно и в составе научного коллектива
(ПК-2)	Способность разрабатывать концептуальные и теоретические модели решаемых научных проблем и задач
<i>2) В проектной и производственно-технологической:</i>	
(ПК-3)	Способность понимания углубленного анализа проблем, постановки и обоснования задач научной и проектно- технологической деятельности
(ПК-4)	Способность разрабатывать концептуальные и теоретические модели решаемых задач проектной и производственно-технологической деятельности
<i>3) В организационно-управленческой деятельности:</i>	
(ПК-5)	Способность управлять проектами, планировать научно-исследовательскую деятельность, анализировать риски, управлять командой проекта
(ПК-6)	Способность разрабатывать и оптимизировать бизнес-планы научно-прикладных проектов
<i>4) В педагогической деятельности:</i>	
(ПК-7)	Способность к преподаванию математических дисциплин и информатики в образовательных организациях основного общего, среднего общего, среднего профессионального и высшего образования
(ПК-8)	Способность разрабатывать учебно-методические комплексы для электронного и мобильного обучения
<i>5) В консалтинговой деятельности:</i>	
(ПК-9)	Способность разрабатывать аналитические обзоры состояния области прикладной математики и информационных технологий
<i>6) В консорциумной деятельности:</i>	
(ПК-10)	Способность к взаимодействию в рамках международных проектов и сетевых сообществ

Реферат

Выпускная квалификационная работа: 92 листа, 7 рисунков, 23 таблиц, 2 приложения, 35 использованных литературных источников.

Ключевые слова: чат-бот, обработка естественного языка, расстояние Дамерау-Левенштейна.

Объект исследования: чат-бот.

Цель работы: разработать прототип чат-бота для автоматизации рутинных задач, с поддержкой естественного языка.

Работа включает: введение, 6 разделов и заключение, список публикаций студента и список использованных литературных источников.

В разделе 1 «Обзор литературы» приведен краткий обзор литературы, необходимый для дальнейшей работы, сформулирована цель и задачи работы.

В разделе 2 «Чат-бот как средство автоматизации работы сотрудников» описаны основные проблемы корпоративных систем, а также приведены возможные пути решения указанных проблем с использованием чат-ботов.

В разделе 3 «Проектирование архитектуры системы» проводится анализ существующих инструментов для обработки естественного языка, рассмотрены инструменты для достижения решений поставленных задач.

В разделе 4 представлены результаты разработки чат-бота.

В разделе 5 «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» рассматривается экономическая составляющая работы, описываются затраты и конкуренты на рынке.

В разделе 6 «Социальная ответственность» оценены риски и опасности для здоровья и окружающей среды.

В заключении приведены основные итоги по работе.

Определения

В данной работе применены следующие термины с соответствующими определениями:

Чат-бот: программа, имитирующая речевое поведение человека при общении с одним или несколькими собеседниками.

Обработка естественного языка: одно из направлений искусственного интеллекта и математической лингвистики, которое занимается изучением проблем компьютерного анализа и синтеза естественных языков.

Естественный язык: это хранящаяся в сознании человека сложная система правил, в соответствии с которыми происходит речевая деятельность, т.е. порождение и понимание текстов.

Искусственный интеллект: область компьютерной науки (информатики), специализирующаяся на моделировании интеллектуальных и сенсорных способностей человека с помощью вычислительных устройств.

Компьютерная лингвистика: направление искусственного интеллекта, которое ставит своей целью использование математических моделей для описания естественных языков.

Базы данных: это информационная модель, позволяющая упорядоченно хранить данные о группе объектов, обладающих одинаковым набором свойств.

Система управления базами данных (СУБД): программное обеспечение, предназначенное для работы с базами данных. СУБД используются для упорядоченного хранения и обработки больших объемов информации.

Нечеткий поиск: поиск информации, при котором выполняется сопоставление информации заданному образцу поиска или близкому к этому образцу значению.

Сервер (аппаратное обеспечение): специализированный компьютер или специализированное оборудование для выполнения на нем сервисного программного обеспечения.

Оглавление

Введение	13
1 Обзор литературы	15
2 Чат-бот как средство автоматизации работы сотрудников	17
2.1 Понятие чат-бота. Этапы развития	17
2.2 Классификация чат-ботов.....	18
2.3 Применение чат-ботов в корпоративных системах.	19
2.4 Корпоративный сайт или чат-бот.....	22
3 Проектирование архитектуры системы	24
3.1 Обработка естественного языка	24
3.2 Компьютерная лингвистика. Анализ текста.....	25
3.3 Этапы обработки естественного языка.....	27
3.3.1 Алгоритмы нечеткого поиска	29
3.4 Разработка чат-бота	32
3.4.1 Архитектура	32
3.4.2 Интерфейс взаимодействия.....	33
3.4.3 Сервер.....	34
4 Разработка и тестирование чат-бота	36
5 Социальная ответственность.....	40
5.1 Правовые и организационные работы обеспечения безопасности	41
5.1.1 Эргономические требования к рабочему месту	41
5.1.2 Особенности законодательного регулирования проектных решений	42
5.2 Производственная безопасность	42
5.2.1 Отклонение показателей микроклимата в помещении.....	43
5.2.2 Недостаток естественного освещения	45
5.2.3 Повышенная напряженность электромагнитного поля.....	46
5.2.4 Недостаточная освещенность рабочей зоны	47
5.2.5 Повышенный уровень электромагнитных излучений	49
5.3 Экологическая безопасность	51
5.4 Безопасность в чрезвычайных ситуациях.....	52

5.5	Выводы по разделу «Социальная ответственность».....	53
6.	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение ...	55
6.1	Предпроектный анализ	55
6.1.1	Потенциальные потребители результатов исследований.....	55
6.1.2	Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения.....	56
6.1.3	SWOT-анализ	57
6.1.4	Оценка готовности проекта к коммерциализации	59
6.1.5	Методы коммерциализации результатов научно-технического исследования	62
6.2	Инициализация проекта.....	63
6.2.1	Цели и результат проекта	63
6.2.2	Организационная структура проекта.....	64
6.3	Планирование управления научно-техническим проектом.....	65
6.3.1	План проекта	65
6.3.2	Бюджет научно-технического исследования (НТИ).....	68
6.4	Реестр рисков проекта	72
6.5	Оценка сравнительной эффективности исследования.....	74
	Выводы по разделу «Финансовый менеджмент».....	75
	Заключение	76
	Список публикаций.....	77
	Список использованных источников	78
	Приложение А	82
	Приложение Б.....	93

Введение

Большинство сотрудников крупных компаний при использовании корпоративного сайта, портала сталкиваются с рядом проблем, в частности

- большая нагрузка на сайт,
- трата времени на выполнение рутинных задач,
- необходимость обучения по использованию корпоративного сайта.

Если для компании ее сайт является не просто визитной карточкой, а средством распространения информации для сотрудников, и включает в себя личные кабинеты, персональные страницы, хранение документов и прочее, то избежать создания сложнейших сайтов просто невозможно. Это приводит к проблемам пользования этими сайтами. Например, портал Томского политехнического университета является достаточно сложным в плане поиска информации. Зачастую проще и быстрее искать информацию на этом сайте с помощью сторонних поисковых сервисов (google, yandex).

Помимо этого, сотрудникам часто приходится тратить время на рутинные задачи, такие как поиск документов, информации о сотрудниках, заполнение документов. В совокупности с загруженностью сайта это лишняя трата рабочего времени. К тому же сотрудников и пользователей необходимо научить пользоваться корпоративными сайтами, а это также дополнительная трата времени. Например, в ТПУ выделяют четыре пары на каждую группу, для того чтобы научить пользоваться сайтом НТБ и искать необходимую литературу. Это без учета того времени, сколько тратит студент на первом курсе чтобы познакомиться с личным кабинетом и информационной составляющей сайта.

Поэтому необходим новый инструмент, который позволит быстрее и проще решать повседневные задачи.

Целью данной работы является разработка прототипа чат-бота для интеграции в информационную систему предприятия.

Для достижения данной цели выделены следующие задачи:

– провести анализ существующих технологий и составить алгоритм работы чат-бота;

– составить алгоритм обработки естественного языка;

– разработать библиотеку для обработки опечаток текста;

– разработать консольный прототип чат-бота;

– провести тестирование компонент чат-бота.

Объектом исследования в данной работе является чат-бот, при этом предмет исследования – Natural language processing (NLP, обработка естественного языка).

1 Обзор литературы

В статье [1] рассмотрены основные проблемы обработки естественного языка. Проанализированы основные направления обработки, методы, инструменты и библиотеки, доступные на текущий момент времени. Проведено два эксперимента, где данные методики были применены для решения реальных задач – анализ тональности новостного фона некоторых криптовалют с целью определения зависимостей между новостным фоном и их обменным курсом и анализ факторов о сотрудничестве компаний, используя их упоминания в различных пресс-релизах. Показано, что анализ текстовых данных имеет большое практическое значение в современном мире.

В работе [2] представлен алгоритм, состоящий из двух этапов – сканирования и верификации. На этапе сканирования осуществляется фильтрация текста с целью исключения из рассмотрения сообщений, точно не содержащих искомые слова. Оставшиеся сообщения проверяются на этапе верификации. Интеграция в поиск лингвистических правил и статистики ошибок позволяет при этом сохранить достаточную точность. В статье приведены оценки эффективности нечеткого поиска в целом, а также используемого в алгоритме классификатора. Тестирование проводилось на выборке текстов сообщений Генерального интернет-корпуса русского языка.

Статья [3] описывает процесс обработки ошибок и замены слов. С коррекциями слов широко сталкиваются в различных приложениях (google, Word и другие). При этом они предоставляют самые вероятные замены для слов, которые, очевидно являются неправильными. В работе описываются алгоритмы коррекции, начиная с расстояния Левенштейна с использованием байесовской оценки решений. Описанный алгоритм проверен на наборе слов из новостных статей «New-York Times». В результате автор делает вывод о том, что теоретическая составляющая метода и результаты испытаний показывают, что алгоритм подходит для проверки орфографии.

Статья [4] описывает подход к разработке виртуальных помощников или чат-ботов в организациях. Авторы создали чат-бот Leo, для поддержки новых студентов во время приемной комиссии. Поэтому в статье описывается системный подход, которым руководствовалась команда при разработке Leo. Основной упор сделан на описание архитектуры, а так же этапов построение рабочей модели ботов.

В работе [5] описывается процесс создания интеллектуального помощника для банков. Основной задачей разработки такого бота является консультация и помощь клиентам. Авторы описывают процесс подбора, обучающий выборки, а так же методы классификации этих данных. Для каждого метода проведены тесты, в результате которых выбрана лучшая модель для последующего использования. Результатом работы является демонстрация обработки запросов пользователей в готовом боте.

В работе [6] представлена методология расчета семантического сходства между двумя словами, предложениями или абзацами. Авторы предлагают использовать лексическую базу данных для сравнения соответствующих значений слов. Для каждого предложения формируется семантический вектор, содержащий вес для каждого слова. Данные веса сравниваются со вторым предложением. Исходя из этого, делается предположение о сходстве или отличии предложений. Как итог, приводятся результаты тестов методологии.

В связи с развитием новых технологий, таких как Siri, Alexa, Cortana и пр. основанных на обработке естественного языка, а так же ростом числа различных мессенджеров, требуется некая стандартизация архитектуры ботов. Автор в статье [7] предлагает и описывает общую архитектуру для проектирования и реализации пользовательских решений чат-ботов. В работе представлено описание каждого этапа разработки бота, а также связи между каждым блоком программы. Автор называет предложенную архитектуру, как масштабируемую и устойчивую.

2 Чат-бот как средство автоматизации работы сотрудников

2.1 Понятие чат-бота. Этапы развития

Чат-бот (chatbot) – программа, целью которой является общение с человеком на понятном ему языке. Под языком может пониматься как естественный язык, так и заготовленные команды и шаблоны.

Впервые понятие чат-бота было использовано в 1994 году Майклом Молдингом. Он создал программу под названием Julia, которая поддерживала общение на естественном языке. Кроме общения, программа также могла самообучаться.

Но первые прототипы чат-ботов были разработаны задолго до этого. В 1964 году профессор Массачусетского института технологий Джозеф Вайнбаум создал программу ELIZA, ставшей первой чат-программой. Ее задачей была имитация общения пользователя с психотерапевтом. Чуть позже была изобретена программа Parry, в которой бот имитировал речь психически больного человека и выдавал себя за пациента. 48% профессиональных психиатров не могли догадаться, что общаются с машиной.

В конце 90-х, начале 2000-х годов начали распространяться чат-боты на платформах IM и IRC (Internet Relay Chat). Как правило, они использовались для модерирования чатов или поддержания беседы.

Более продвинутые боты появились позднее. В 2006 году появилась программа Watson. Ее особенностью являлась способность распознавать не только текст, но и голосовые сообщения. Данная технология послужила основой для создания многих голосовых помощников.

С 2010 года крупные компании начали использовать и внедрять своих ботов: Siri (компания Apple), Google Now (компания Google). Так же свои наработки представили Amazon и Microsoft.

Главной задачей чат-ботов является выполнение запросов пользователей. Поддержание беседы на сегодняшний день не представляет интерес для массового потребителя. Но кроме грамотного выполнения задач,

важным критерием является удобство пользования этим инструментом. Поэтому следующим этапом развития ботов стало активное распространение мессенджеров. По статистике число пользователей WhatsApp в 2016 году достигло 1,6 миллиард пользователей. В тоже время в тройку популярных мессенджеров в России вошел “Вконтакте”. На основе этих мессенджеров строятся диалоговые боты для абсолютно разных целей: информационные задачи (рассылка новостей, поиск информации), развлечения (интерактивные игры, юмор), продажи (консультирование, оформление заказов) и пр.

Первой компанией, внедрившей коммерческого чат-бота в свой мессенджер, стала компания Facebook. С этих пор развитие платформы ботов позволило компаниям создавать свои инструменты для привлечения клиентов и автоматизации многих задач.

2.2 Классификация чат-ботов

Глобально ботов делят на два типа:

- боты, принимающие определенные команды;
- боты, способные анализировать поступающую информацию и извлекать смысл из текста.

Первый тип ботов, как правило, используется для узконаправленных целей и не рассчитан на массового потребителя. Для работоспособности в его алгоритмы заранее закладываются определенные команды. Тем не менее, до сих пор большинство ботов работают по данному принципу.

Ко второй категории относятся более сложные программы. Данные боты, как правило, используют нейронные сети и глубокое обучение, в связи с чем, способны принимать любые запросы пользователя, обрабатывать их и выдавать ответы. Таким образом, бот способен общаться на естественном для человека языке.

Под нейронными сетями подразумеваются математические модели, а также их программные или аппаратные реализации, построенные по принципу

организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма.

Обработка естественного языка (Natural Language Processing, NLP) - обобщенное направление искусственного интеллекта, а также математической лингвистики. Здесь изучаются проблемы синтеза естественного языка и компьютерного анализа.

2.3 Применение чат-ботов в корпоративных системах.

Ценностью чат-ботов является возможность избавить пользователя от использования нагруженных интерфейсов, позволяя привычно взаимодействовать с субъектами коммуникации в интерактивном режиме.

Боту можно задать вопрос и тут же получить на него ответ, что избавляет от необходимости самостоятельного поиска этого ответа в массивах информации. В этом смысле чат-бот подобен информационно-поисковой системе, однако ценность его может быть ещё выше. Поскольку процесс коммуникации с чат-ботом интерактивный, как с живым человеком, это позволяет заинтересовать пользователя в продолжении коммуникации и повысить степень его вовлечения.

На сегодняшний день крупные компании ищут пути внедрения интеллектуальных помощников, которые способны выполнять различные рутинные офисные задачи. Среди таких задач можно выделить:

- подбор персонала;
- адаптация и обучение новых сотрудников;
- ответы на вопросы;
- оформление документов;
- экзит-интервью;
- планирование времени и ресурсов и пр.

Роботизация бизнес-процессов значительно ускоряет процессы за счет автоматизации пользовательских интерфейсов. Робот способен выполнять свои функции и задачи непрерывно, что дает явное преимущество для ускорения

бизнес-процессов с более быстрой обработкой запросов и улучшением качества обслуживания клиентов.

Сложные корпоративные системы несут в себе некоторые проблемы. Необходимо создавать нагромождённые пользовательские интерфейсы, для использования которых сотрудникам необходимо предварительно обучаться. Такие системы включают в себя хранение документации (ГОСТы, статьи, приказы), корпоративную почту, личные кабинеты с различными функциями. К тому же зачастую, разработчики не предусматривают адаптацию систем под мобильные платформы, что создает неудобство для сотрудников, находящихся за пределами офиса.

Поэтому можно выделить первостепенный функционал для чат-ботов:

- предоставление информации (номера и электронные почты сотрудников, плановые мероприятия);
- оповещение сотрудников компании о необходимых новостях;
- рассылка напоминаний;
- поиск информации и документов, как во внутрикорпоративных сетях, так и в интернете;
- отправка электронных сообщений;
- заполнение документов.

Использование заготовленных команд для такого количества функций не является удобным решением. В связи с чем, можно выделить в качестве функционала способность бота общаться на естественном для человека языке.

Для ценности бота среди персонала, необходимо выделить ряд правил:

- чем меньше действий бот требует от пользователя, тем лучше.

Ценность внедрения интеллектуального помощника, как раз и заключается в автоматизации рутинных процессов.

- сообщения от бота должны быть информативными и краткими, так как большое количество информации может заставить пользователя искать объяснения в переписке выше.

– бот должен активно предугадывать потребности пользователя. Необходимо не только правильно выполнять задачи, но и искать самые оптимальные решения.

Для любого крупного предприятия первостепенным критерием внедрения нового инструмента в корпоративные сети является его безопасность. Под безопасностью подразумевается:

- отсутствие возможности утечки корпоративных данных;
- отсутствие угрозы взлома внутрикорпоративной сети через этот инструмент;
- надежность инструмента.

Критерии к администраторам платформы чат-бота:

- наличие начальных знаний в программировании и системном администрировании;
- знания в работе с базами данных;
- наличие опыта в интеграции программных продуктов и платформ;
- наличие опыта в постановке задач по доработке и улучшению внедряемых программных продуктов и платформ.

Серверное оборудование и все компоненты платформы для чат-бота должны работать без сбоев круглосуточно и ежедневно за исключением случаев, когда необходимо произвести регламентные или профилактические работы на серверном оборудовании, либо предстоят технические работы, в результате которых произойдет отключение электроснабжения серверного оборудования.

Для устойчивой работы чат-бота, платформа должна располагаться на выделенном сервере. Данный сервер должен работать круглосуточно и ежедневно без сбоев. Отключение должно происходить только во время регламентных или профилактических работ.

В платформе должна быть предусмотрена функция мягкой остановки, в результате которой есть возможность безопасно закрыть все активные соединения, завершить все службы, сохранить информацию и перейти в режим

ожидания. После окончания всех регламентных работ, платформа должна возобновить свою работу, в результате которой восстанавливаются все службы и компоненты.

Защита информации на платформе, которая обрабатывает данные от чат-бота должна быть реализована как на уровне доступа к данным, так и на уровне передачи информации по каналам связи.

2.4 Корпоративный сайт или чат-бот

Выделим основные задачи, которые решает корпоративный портал.

1. Организация внутренних коммуникаций. Порталы сравнивают с социальными сетями, где сотрудники могут общаться в привычном формате. Наиболее популярная функция – «живая лента», где люди получают информацию в виде постов.

Также порталы предлагают свои чаты, которые должны заменить мессенджеры. Однако их сложно назвать полноценной заменой. Люди привыкли к своим мессенджерам, которые всегда под рукой — в телефоне или отдельном приложении на рабочем компьютере (а не на перегруженном веб-портале) — и в результате используют именно их.

2. Быстрый поиск информации и документов. Корпоративные порталы решают проблему разрозненного хранилища документов. Также они помогают с онбордингом (адаптацией) новых сотрудников, предоставляя быстрый доступ к нужной информации. Но удобно ли искать нужный документ (например, шаблон заявления на отпуск) среди сотен других?

3. Автоматизация внутренних бизнес-процессов. Постановка и трекинг выполнения задач, отправка заявлений на отгулы и отпуска — всё это можно выполнять в типовом корпоративном портале. И если речь идёт о сложных бизнес-процессах, то без портала точно не обойтись.

Кроме того, таск-трекинг там обычно дополняется удобными календарями и возможностью брони переговорных.

Интеграции с другими внутренними системами.

Все разработчики порталов обещают возможность интеграции с другими системами компании. И это логично, сейчас все говорят об экосистемах, и мало кто может существовать только с одним инструментом, решая все свои задачи в нём.

Приведем краткое описание как чат-боты решают указанные выше задачи.

1. *Организация внутренних коммуникаций.* Чат-бот устанавливается именно в тот мессенджер, где сотрудники общаются друг с другом и решают рабочие задачи. Таким образом, бот естественно встраивается в рабочую коммуникационную среду, становясь ещё одним членом коллектива (виртуальным).

Информационное сообщение, отправленное от лица чатбота, становится более персонализированным, чем простая новость на портале. А значит, сотрудник обратит на него больше внимания (как минимум, прочитает).

По данным StackOverflow 87% сообщений в мессенджере могут быть прочитаны в течение 3 минут с момента отправки. Корпоративный портал или email вряд ли смогут дать сопоставимые показатели.

2. *Быстрый поиск информации и документов.* Чат-бот позволяет искать информацию или документы в заданных источниках. Для этого достаточно в относительно свободной форме выполнить запрос с указанием данных, которые необходимо найти.

Данные ограничиваются только настройками самого бота. То есть поиск информации может выполняться только в доступных ему местах.

Чатбот готов мгновенно выдавать любую информацию: от дня рождения или контакта коллеги до пароля к Wi-Fi. При этом бот может не только отвечать на вопрос в свободной форме, но и позволяет самостоятельно найти информацию по дереву категорий. Такая функция будет полезна новым сотрудникам, которые пока не понимают, какой именно вопрос задать. Они могут посмотреть все категории и ознакомиться со всей доступной

информацией. Тем более, многие в начале своей работы в компании стесняются задавать вопросы коллегам, а спросить у бота – не страшно.

3. *Автоматизация внутренних бизнес-процессов.* Если стоит задача автоматизации сложных бизнес-процессов, то не обойтись без привычного веб-интерфейса. Но существуют задачи, для которых нет необходимости использовать корпоративные сайты, например, заполнение и отправка заявления на отпуск. Для этого достаточно выполнить запрос у бота, после чего, ввести для него требуемую информацию и ждать оповещения о результатах рассмотрения. При этом нет необходимости обращаться лично в соответствующие отделы, и заполнять документы собственноручно.

4. *Интеграции с другими внутренними системами.* Чат-бот может быть интегрирован с любой системой не хуже корпоративного портала.

Вся логика работы бота реализуется на сервере, поэтому ему доступны все возможности любого программного продукта. Мессенджер — это лишь интерфейс, с помощью которого бот общается с пользователем. Поэтому за кажущейся простотой скрывается довольно мощный инструмент.

3 Проектирование архитектуры системы

3.1 Обработка естественного языка

Компьютеры способны качественно и быстро обрабатывать любые структурированные данные. В естественном же языке отсутствует строгая структуризация и типизация. Это приводит к тому, что компьютерам сложно обрабатывать поступающую информацию, если она не соответствует строгим правилам.

В каждом языке можно выделить свои правила и структуры, но этого не достаточно для качественной обработки. К тому же, если рассматривать тексты, составленные обычными пользователями интернета, то можно найти огромное количество ошибок, как синтаксических, так и орфографических.

Все это приводит к проблемам обработки естественного языка. Этими проблемами занимается отдельное направление в искусственном интеллекте – обработка естественного языка (Natural Language Processing, NLP).

3.2 Компьютерная лингвистика. Анализ текста

Для полноценной обработки текста необходимо проводить его полный анализ. При этом до сих пор не все теоретические наработки реализованы на практике, что приводит к определенным трудностям. Основной проблемой является сложность синтаксического анализа текста. Сложность заключается в том, что естественный язык служит не только для формальной обработки информации, но и для живого воображения. Зачастую, анализ объемных текстов, является сложной задачей не только для ЭВМ, но и для человека, так как большинство диалогов имеют под собой какую-то реальную или вымышленную основу.

Для полноценного анализа текста, готовая система должна анализировать текст с точки зрения синтаксиса (структуры предложения), семантики (понятий, применяемых в тексте) и прагматики (правильности употребления понятий и целей их употребления). После анализа, система должна сгенерировать свой ответ, пригодный для логического вывода.

Полноценная система, с возможностью полного анализа, должна иметь следующие модули.

Графематический анализ – это система первостепенного анализа естественного текста, подготавливающая информацию для последующей обработки (морфологический и синтаксический анализ). Как правило, графематический анализ разбивает исходный текст на различные составляющие: выделение абзацев и заголовков, разделение текста на слова, выделение фамилии имени отчества (в том, случае, если имя и отчество написаны инициалами), электронных адресов, выделение предложений, выделение устойчивых оборотов и пр.

Морфологический анализ – позволяет определять нормальную форму слов. Это необходимо для того, что бы обрабатывать текст исходя из нормальной формы, а не различных словоформ.

Предсинтаксический анализ – служит для объединения отдельных лексических единиц в одну синтаксическую или, наоборот, ее разделение на несколько. В одну синтаксическую единицу объединяются изменяемые неразрывные словосочетания (например, «бить баклуши»). Делением слов особенно необходимо заниматься, например, в немецком языке, где несколько произвольных связанных между собой слов могут объединяться в одно сложное «на лету», а помещать в морфологический анализ все подобные сочетания не представляется возможным.

Синтаксический анализ – самая сложная часть анализа текста. Здесь необходимо определить роли слов и их связи между собой. Результатом этого этапа является набор деревьев, показывающих такие связи. Выполнение задачи осложняется огромным количеством альтернативных вариантов, возникающих в ходе разбора, связанных как с многозначностью входных данных (одна и та же словоформа может быть получена от различных нормальных форм), так и неоднозначностью самих правил разбора.

Постсинтаксический анализ служит двум целям. С одной стороны необходимо уточнить смысл, заложенный в слова и выраженный при помощи различных средств языка: предлогов, префиксов или аффиксов, создающих ту или иную словоформу. С другой стороны, одна и та же мысль может быть выражена различными конструкциями языка.

Семантический анализ проводит анализ текста «по смыслу». С одной стороны, семантический анализ уточняет связи, которые не смог уточнить постсинтаксический анализ, так как многие роли выражаются не только при помощи средств языка, но и с учетом значения слова. С другой стороны, семантический анализ позволяет отфильтровать некоторые значения слов или даже целые варианты разбора как «семантически несвязные».

3.3 Этапы обработки естественного языка

Компьютерная лингвистика — раздел науки, изучающий применение математических моделей для описания лингвистических закономерностей. Ее можно разделить на две большие части. Одна из них изучает способы применения вычислительной техники в лингвистических исследованиях — применение известных математических методов (например, статистическая обработка) для выявления закономерностей. Обнаруженные закономерности используются другой частью, изучающей вопросы осмысления текстов, написанных на естественном языке, — создание математических моделей для решения лингвистических задач и разработка программ, функционирующих на основе этих моделей. Эта часть компьютерной лингвистики тесно соприкасается с разделом искусственного интеллекта, занимающегося разработкой систем обработки текстов на естественном языке.

Общая схема обработки текстов (рис. 1) инвариантна по отношению к выбору естественного языка. Независимо от того, на каком языке написан исходный текст, его анализ проходит одни и те же стадии. Первые две стадии (разбиение текста на отдельные предложения и на слова) практически одинаковы для большинства естественных языков. Единственное, где могут проявиться специфичные для выбранного языка черты, — это обработка сокращений слов и обработка знаков препинания (точнее, определение того, какие из знаков препинания являются концом предложения, а какие нет).

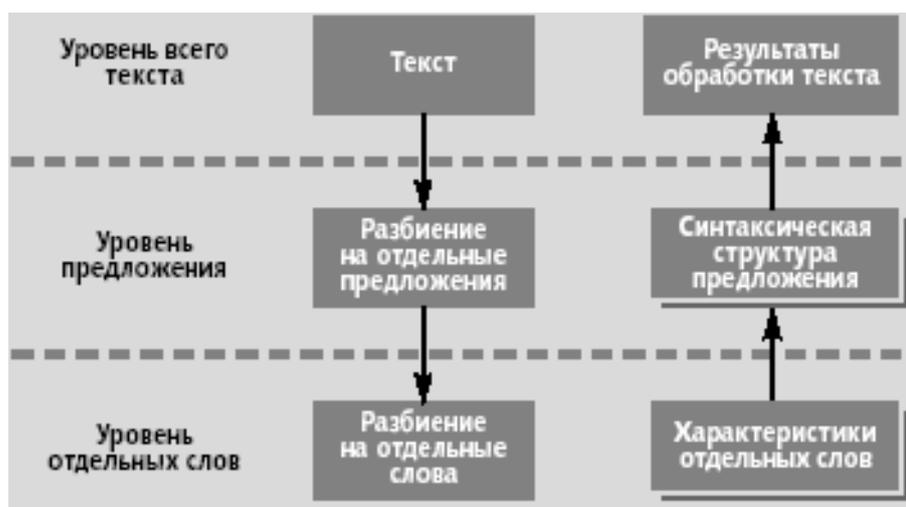


Рисунок 1 – Этапы обработки естественного языка

Последующие две стадии (определение характеристик отдельных слов и синтаксический анализ), напротив, сильно зависят от выбранного естественного языка. Последняя стадия (семантический анализ) также мало зависит от выбранного языка, но это проявляется только в общих подходах к проведению анализа.

Семантический анализ основывается на результатах работы предыдущих фаз обработки текста, которые всегда специфичны для конкретного языка. Следовательно, способы представления их результатов тоже могут сильно варьироваться, оказывая большое влияние на реализацию методов семантического анализа. Результаты анализа, произведенного на ранних стадиях, могут быть многозначны: для выходных параметров указывается не одно, а сразу несколько возможных значений (скажем, может существовать несколько способов трактовки одного и того же слова). В таких случаях последующие стадии должны выбирать наиболее вероятные значения результатов ранних стадий анализа и уже на их основе проводить дальнейший анализ текста.

3.3.1 Алгоритмы нечеткого поиска

Большой проблемой при наборе текста в мессенджерах являются опечатки. По статистике в среднем в запросе 3 слова. При этом 12% запросов с опечатками, 84% из которых с одной ошибкой.

Типы опечаток:

1. Ошибки в словах:
 - 1.1. Замена буквы (афтомобиль → автомобиль);
 - 1.2. Пропуск буквы (бмага → бумага);
 - 1.3. Вставка буквы (клавииаатура → клавиатура);
 - 1.4. Перестановка букв (трно → трон);
2. Склейка и разрезание:
 - 2.1. Пропуск пробела (крышадом → крыша дома);
 - 2.2. Вставка пробела (пол день → полдень);
3. Раскладка клавиатуры (ghbdt → привет).

Пусть имеется две строки a и b , требуется определить их сходство. На этапе разработки прототипа программы будем рассматривать только синтаксическое сходство, то есть сравнение последовательности символов, для поиска опечаток.

Для алгоритма нечеткого поиска выделим две меры: а) точное совпадение, б) преобразование. Мера *точного совпадения* выдает один из двух возможных результатов: 1 – строки одинаковы, 0 – строки отличаются. Мера преобразования ведет подсчет количества операций, необходимых для превращения одной строки в другую. Наиболее популярными методами подсчета количества данных операций являются методы поиска редакционного расстояния, которое измеряет минимальное количество операций редактирования, необходимых для преобразования строки a в строку b .

Среди методов поиска редакционного расстояния можно выделить: расстояние Левенштейна, Дамерау-Левенштейна, алгоритм Нидлмана-Вунша, Смита-Ватермана, Смита-Ватермана-Гота, расстояние Хэмминга. В работе [10] приведено сравнение вышеперечисленных алгоритмов по скорости работы, а

так же точности сравнения слов. Согласно результатам исследования самым точным является расстояние Хэмминга. Но данный алгоритм рассматривает только строки одинаковой длины, что накладывает ограничения на возможные варианты замены. Оптимальным вариантом является расстояние Дамерау-Левенштейна, так скорость данного алгоритма является одной из самых высоких среди конкурентов, при этом точность поиска расстояния высока.

Расстояние Дамерау-Левенштейна

Определение. Расстояние Дамерау-Левенштейна между двумя строками называется минимальное количество операций вставки, удаления, замены и транспозиции (перестановки двух соседних символов), необходимых для перевода одной строки в другую. Является модификацией расстояния Левенштейна, отличается от него добавлением операции вставки.

Расстояние Дамерау-Левенштейна между двумя строками a и b определяется функцией $d_{a,b}(|a|, |b|)$ как:

$$d_{a,b}(i, j) = \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0, \\ \min \begin{cases} d_{a,b}(i-1, j) + 1 \\ d_{a,b}(i, j-1) + 1 \\ d_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{if } i, j > 1 \text{ and } a_i = b_{j-1} \text{ and } a_{i-1} = b_j \\ \min \begin{cases} d_{a,b}(i-1, j) + 1 \\ d_{a,b}(i, j-1) + 1 \\ d_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise,} \end{cases} \quad (1)$$

где $1_{(a_i \neq b_j)}$ это индикаторная функция, равная нулю при $a_i = b_j$ и 1 в противном случае, $|a|$ – длина строки a .

Каждый рекурсивный вызов соответствует одному из случаев:

- $d_{a,b}(i-1, j) + 1$ соответствует удалению символа (из a в b);
- $d_{a,b}(i, j-1) + 1$ соответствует вставке (из a в b);
- $d_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)}$ соответствие или несоответствие, в зависимости от совпадения символа;
- $d_{a,b}(i-2, j-2) + 1$ в случае перестановки двух последовательных символов.

Так как расстояние Дамерау-Левенштейна опирается на расстояние Левенштейна, то поиск по данному алгоритму может вывести несколько кандидатов на замену. Например, «стадо *свией* бросилось в море». Слово *свией* может быть заменено несколькими словами: *сваей*, *свиной*, *своей*, *свей* и пр.

В таком случае необходимо определить: какое слово наиболее вероятно в данном контексте.

Математически постановку задачи можно представить следующим образом:

$$\hat{t} = \operatorname{argmax}_t \rho(t|\omega), \quad (2)$$

где \hat{t} – наилучшее исправление, $t = t_1 \dots t_n$ – возможные исправления, ω – исходное предложение.

По формуле Байеса:

$$\hat{t} = \operatorname{argmax}_t \rho(t|\omega) = \operatorname{argmax}_t \frac{\rho(t|\omega)\rho(t)}{\rho(\omega)}. \quad (3)$$

При условии, что $\rho(\omega)$ всегда одинаково, получаем:

$$\hat{t} = \operatorname{argmax}_t \rho(t|\omega)\rho(t),$$

где $\rho(t) = \rho(t_1 \dots t_n)$ – вероятность увидеть текст $t_1 \dots t_n$,

$\rho(\omega, t) = \rho(\omega_1 \dots \omega_n | t_1 \dots t_n)$ – вероятность, что в результате опечаток текст $t_1 \dots t_n$ будет преобразован в $\omega_1 \dots \omega_n$.

3.4 Разработка чат-бота

3.4.1 Архитектура

Чат-боты – программы участвующие в диалоге с поочередной передачей инициативы, целью которой является интерпретация входного текста или речи и вывод соответствующего полезного ответа. В отличие от людей, с которыми они взаимодействуют, для достижения этих результатов чат-боты должны полагаться на эвристики и методы машинного обучения.

На рисунке 2 представлена общая архитектура чат-бота.



Рисунок 2 – Архитектура чат-бота

Архитектура состоит из двух основных компонентов. Первый — интерфейс с пользователем, осуществляющий прием ввода пользователя (например, через микрофоны для ввода речи или через веб-интерфейс для ввода текста) и передачу интерпретируемого вывода (через динамики для вывода синтезированной речи или через экран мобильного устройства для вывода текста). Этот внешний компонент обертывает второй — внутреннюю диалоговую систему, которая интерпретирует входной текст, управляет внутренним состоянием и генерирует ответы.

Любой чат-бот подразумевает разработку как интерфейсной части (frontend), так и серверной (backend) . К каждой из этих частей выдвигаются определенные условия.

При выборе платформы интерфейса выдвигается ряд правил: распространённость платформы, алгоритм работы, структура взаимодействия, возможности, доступные для использования операционные системы, типы лицензии, сложность внедрения.

Для предприятия в первую очередь важна безопасность приложения, то есть отсутствие утечки информации и корпоративных данных. К тому же, программа разрабатывается для упрощения работы персонала. Поэтому необходимо четко проработать каждый из выше перечисленных требований.

Серверная часть подразумевает, в первую очередь, алгоритмы обработки запросов пользователей. К тому же должна быть проработана адаптация работы бота с уже имеющимися корпоративными сервисами и базами данных. Общая схема работы программы представлена на рисунке 3.

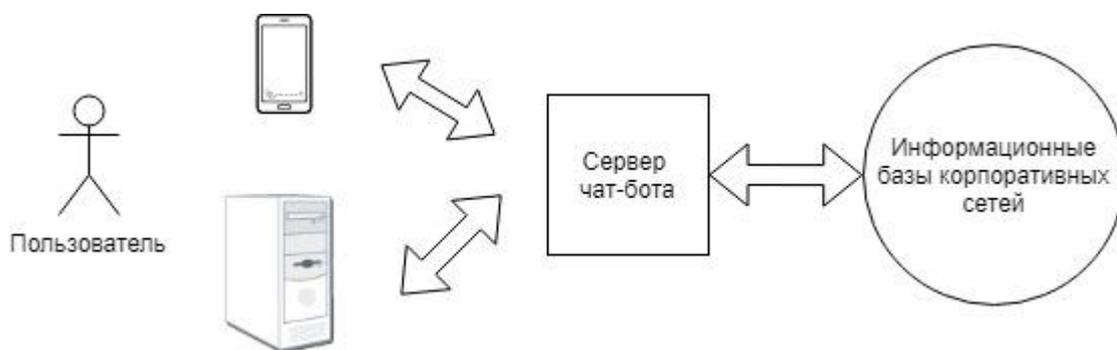


Рисунок 3 – Общая схема чат-бота в корпоративной системе

3.4.2 Интерфейс взаимодействия

Глобально разделить интерфейсную часть можно на две категории: разработка собственного чата и использование уже доступных мессенджеров (например, telegram, Rocket.chat, Slack, Skype).

Чат-бот создается для упрощения взаимодействия пользователя с корпоративными данными и функциями. Поэтому строго привязывать его к

компьютеру, не имеет ни какого смысла. Соответственно, необходимо использовать практически все популярные операционные системы, например, Windows, Linux, IOS, Android. Разработка своего чата под такое количество платформ затянет разработку приложения, а так увеличит конечную стоимость продукта. Поэтому принято решение выбрать один из имеющихся мессенджеров.

Главным критерием является безопасность приложения по отношению к утечке персональных и корпоративных данных. Поэтому использовать распространенные программы, например telegram, не предоставляется возможности. Данные программы работают на сторонних серверах (в том числе и иностранных), протоколы обмена и хранения информации из диалогов зашифрованы.

Таким образом, необходимо выбрать продукт с открытым исходным кодом. Данному критерию удовлетворяет мессенджер Rocket.chat. Это проект с открытым программным кодом. Положительными сторонами данного мессенджера является гибкая настройка и поддержание программы всех актуальных операционных системах, включая web-версию. Устанавливается платформа на сервере компании, что гарантирует отсутствие утечку любой информации.

Стоит отметить, что Rocket.chat необходимо устанавливать на сервер, а так же проводить настройки, касающиеся обмена сообщения. Это создает неудобство в тестировании, а так же дополнительную нагрузку на процессор. Вследствие этого, для разработки и тестирования прототипа можно использовать более простые интерфейсы, например, telegram. Он не требует настройки и установки, что упрощает процесс тестирования. Соответственно, для тестирования был выбран именно этот мессенджер.

3.4.3 Сервер

В качестве основы сервера выбран язык программирования Python, являющийся языком общего направления. Данный язык является наиболее

популярным в работе с обработкой естественного языка и нейронными сетями. Хотя Python и является относительно медленным языком, тяжелые библиотеки разработаны на C, C++ и портированы на Python, что позволяет им работать без потери скорости. Основные библиотеки для обработки естественного языка - rtermextract и rumorphy2.

Для хранения истории запросов пользователей, а так же информации о пользователях необходимо использовать базы данных. В качестве хранилища используется PostgreSQL. PostgreSQL – объектно-реляционная СУБД с открытым исходных кодом. Она поддерживает большую часть стандарта SQL.

В базе данных должна храниться вся переписка пользователей, включая удаленные сообщения. Прототип схемы базы данных для хранения истории запросов и информации о пользователях представлен на рис. 4.

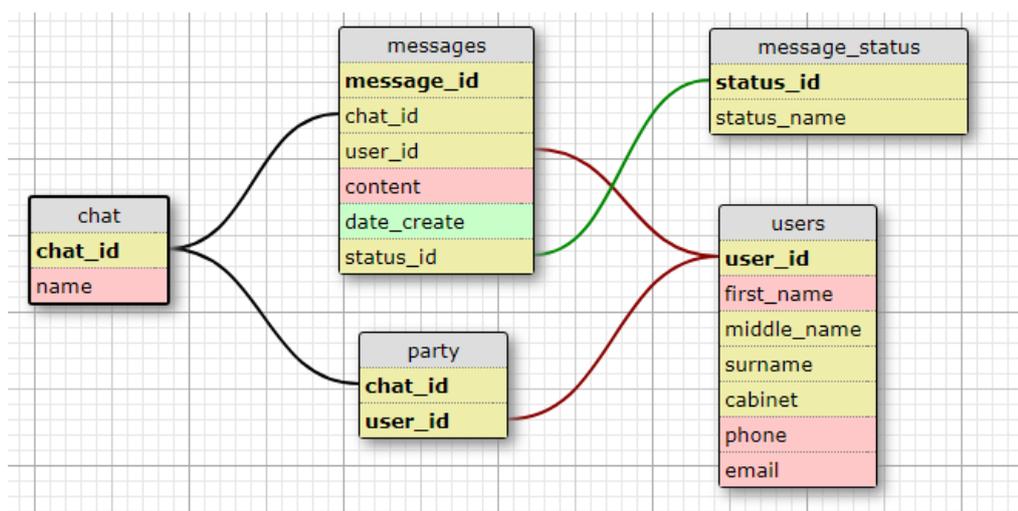


Рисунок 4 – Схема хранения истории запросов и информации о пользователях в прототипе чат-бота

Прототип учитывает историю чата не только с ботом, но также и с другими пользователями (в том числе и групповые чаты). PostgreSQL позволяет получать информацию из других СУБД, например MySQL. Поэтому для тестирования прототипа достаточно создать одну таблицу с основными данными сотрудников. В дальнейшем эта информация будет поступать из уже имеющихся корпоративных баз данных.

4 Разработка и тестирование чат-бота

Сформулируем задачи, которые должен решать прототип чат-бота:

- поиск информации о сотрудниках из базы данных;
- отправка электронных писем;
- поиск информации в корпоративной системе;
- поиск документов в корпоративной системе;
- автоматическое заполнение документов;

Поиск информации о сотрудниках из базы данных.

Для поиска информации о сотрудниках требуется в обязательном порядке ввести фамилию и имя. В таком случае будет выведена полная информация обо всех сотрудниках с такими фамилиями и именами. В случае если пользователь введет необходимый ему параметр сотрудника (телефон, кабинет, электронный адрес), то представится конкретная информация по данным пунктам. Пример запросов и ответов представлен в таблице 1.

Таблица 1. Пример запросов о поиске информации о сотрудниках

C:\Windows\system32\cmd.exe	
<<<	“Антон Сергеевич Иванов”
>>>	Кабинет: 217 Телефон: 2-32-18 Электронный адрес: aai@mail.ru
<<<	“Антон Иванов”
>>>	Антон Сергеевич Иванов Кабинет: 217 Телефон: 2-32-18 Электронный адрес: aai@mail.ru Антон Антонович Иванов Кабинет: 217 Телефон: 2-32-18 Электронный адрес: aai@mail.ru
<<<	“Антон Сергеевич Иванов телефон кабинет”
>>>	Антон Сергеевич Иванов

Телефон: 2-32-18 Кабинет: 217
<<< "бот, пожалуйста, выведи номер телефона, пожалуйста, антона сергеевича иванова"
>>> Антон Сергеевич Иванов Телефон: 2-32-18

Отправка электронных сообщений

Для отправки электронных сообщений необходимо указать ключевое слово «электронное сообщение» или похожее по смыслу. Можно так же указать в этом же сообщении электронные адреса или имена сотрудников. В случае если сотрудников с таким именем несколько, бот выведет список электронных адресов соответствующих введенному имени и попросит пользователя ввести адреса вручную. Отдельным сообщением потребуются ввести текст сообщения и тему. Пример работы бота при отправке сообщений представлен в таблице 2.

Таблица 2. Пример диалога отправки электронного сообщения

C:\Windows\system32\cmd.exe	
<<<	“Отправить электронное письмо”
>>>	Введите адреса или имена получателей
<<<	“gena_chursin1995@mail.ru”
>>>	Введите тему сообщения
<<<	“Тест”
>>>	Введите текст сообщения
>>>	“Пробное сообщение”
<<<	Письмо отправлено

Аналогично, можно отправлять сообщения сразу на несколько адресов. Во избежание неточностей, на данный момент, рекомендуется отправлять сообщения именно по такой структуре. Тем не менее, форма отправки может быть любой. В случае неоднозначности программа запросит уточнение.

Поиск документов в корпоративной системе

Поиск документов в корпоративной системе осуществляется в два этапа. Пользователь запрашивает поиск документов, для этого отправляется сообщение “документ” или подобные формы. Далее бот запрашивает название документа. Название выведено в отдельное сообщение, для того, чтобы

исключить ошибки четкого определения именно названия. В ответ приходит сообщение с документом. Пример запроса представлен на рисунке 4.

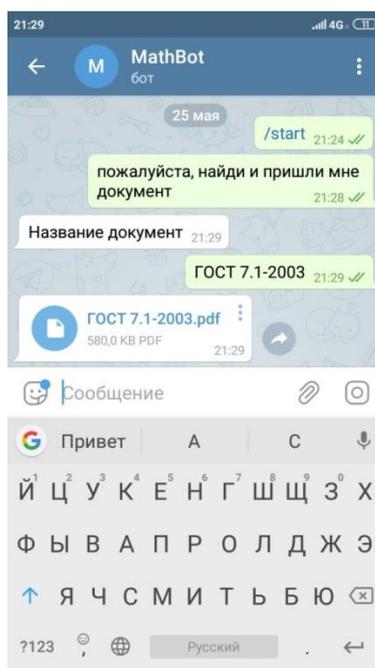


Рисунок 5 – Пример поиска документа в системе

На данный момент алгоритм работает следующим образом. Существует таблица (или в данном случае документ), в котором есть два поля: имя документа и путь к файлу. После получения имени документа, по данному пути извлекается документ и передается в чат.

Проблема заключается в поиске документов. Пользователю необходимо знать точное название документа, хранящееся в базе. В противном случае он не получит его.

Автоматическое заполнение документов

Для автоматического заполнения документов боту необходимо сказать «заполнить заявление» или по теме «заполнить заявление на отпуск», после чего бот запросит несколько данных и выдаст ответ.

Проблема в выполнении задачи состоит в том, что необходимо заранее закладывать в базу те вопросы, которые боту необходимо задать для заполнения бумаг.

Пример заполненного заявления представлен в приложении 1. В данном примере бот задал два вопроса, о начале отпуска и продолжительности. ФИО пользователя и дата взяты из БД.

Обработка опечаток

В качестве словаря априорной вероятности используется подготовленная база, составленная из статей с сайта Wikipedia, а так же новостных сайтов.

Таблица 3. Результаты обработки опечаток

C:\Windows\system32\cmd.exe	
<<<	“Привет”
>>>	Возможные варианты исправления опечаток: «Привет»
<<<	«Стадо свией»
>>>	Возможные варианты исправления опечаток: «свиной» «сваей» Скорее всего, вы имели ввиду «свиной»

Расстояние Дамерау-Левенштейна корректно ищет расстояние между словами. Далее по частотному словарю ищется вероятная замена. Примеры ответов при опечатке с одним и двумя кандидатами на замену представлены в таблице 3.

Слабым местом в реализованном алгоритме, является скорость поиска замен. На данный момент, расчеты занимают около 3 секунд. Первой причиной является компьютер, на котором проводились тесты. Данный процессор не подходит для скоростной обработки больших объемов информации. Использование профессиональных серверов, позволит сократить данное время более чем в два раза. Второй причиной является словарь, в котором представлено более 120000 слов. Решение данной проблемы – удаление кандидатов, которые точно не нужны для запросов в данной программе. Но для этого, необходимо проводить тестирование с реальными пользователями и собирать данные для анализа.

5 Социальная ответственность

В данном разделе указаны такие вредные факторы, оказывающие негативное влияние на организм человека, как электромагнитное излучение, неоптимальный микроклимат помещения, недостаточность освещения и такой опасный фактор, как электрический ток. Так же указан характер вредного воздействия данных факторов на организм и последствия их длительного и чрезмерного воздействия.

Так же рассмотрены вопросы, связанные с организацией рабочего места и условий в которых будет реализовываться разработка, полученная в ходе написания магистерской диссертации, а именно, обработка данных, полученных на рабочем месте, в соответствии с нормами производственной санитарии, техники безопасности и охраны труда и окружающей среды.

Условия труда оказывают прямое воздействие на здоровье человека и его состояние в процессе работы, что обуславливает их отвечать всем требованиям безопасности и санитарно-гигиеническим требованиям. Все факторы, воздействующие на работника в процессе осуществления трудовой деятельности, принято разделять на два типа: вредные и опасные производственные факторы.

Рабочим местом является аудитория 427а 10-ого корпуса Томского политехнического университета по адресу проспект Ленина 2, рабочей зоной является стол с компьютером.

В результате постоянной физической или моральной нагрузки, испытывая сильное напряжение, возникает хронический стресс. В свою очередь, стресс способен повлиять на жизнедеятельность организма, то есть, ухудшить работу органов дыхания, сердечно-сосудистой системы, пищеварения, желез внутренней секреции. Данная проблема часто рассматривается в рамках психофизиологии.

Далее были указаны ЧС, которые могут произойти на рабочем месте и действия, которые необходимо выполнить в случае их возникновения.

5.1 Правовые и организационные работы обеспечения безопасности

5.1.1 Эргономические требования к рабочему месту

В процессе работы, все используемые предметы должны находиться в зоне досягаемости. Оптимальное размещение предметов труда и документации в зонах досягаемости приведено на рисунке 4.1.

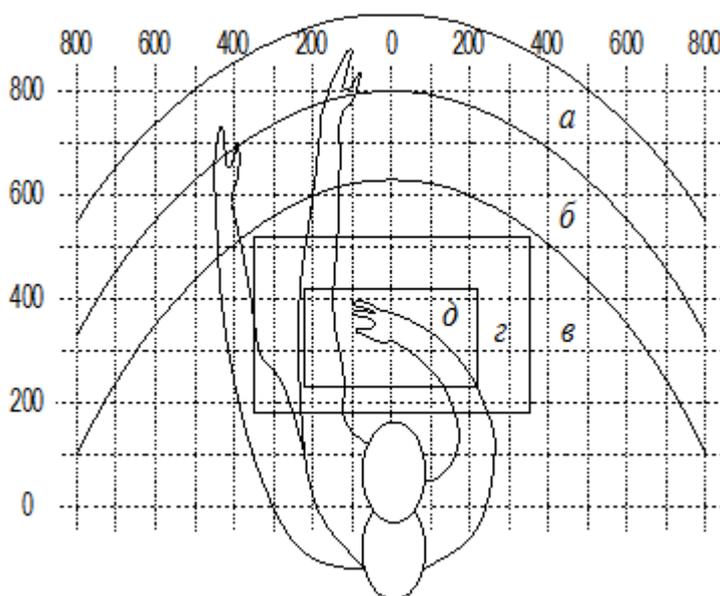


Рисунок 6 – Оптимальное размещение предметов труда и документации в зонах досягаемости

На рисунке 6 введены следующие обозначения:

- а – зона максимальной досягаемости;
- б – зона досягаемости пальцев при вытянутой руке;
- в – зона легкой досягаемости ладони;
- г – оптимальное пространство для грубой ручной работы;
- д – оптимальное пространство для тонкой ручной работы [10]

5.1.2 Особенности законодательного регулирования проектных решений

В соответствии с государственными стандартами и правовыми нормами обеспечения безопасности предусмотрена рациональная организация труда в течение смены, которая предусматривает:

- длительность рабочей смены не более 8 часов;
- установление двух регламентируемых перерывов (не менее 20 минут после 1-2 часов работы, не менее 30 минут после 2 часов работы);
- обеденный перерыв не менее 40 минут.

Обязательно предусмотрен предварительный медосмотр при приеме на работу и периодические медосмотры.

Каждый сотрудник должен пройти инструктаж по технике безопасности перед приемом на работу и в дальнейшем, должен быть пройден инструктаж по электробезопасности и охране труда.

Предприятие обеспечивает рабочий персонал всеми необходимыми средствами индивидуальной защиты. Таким образом, по итогу проделанной работы можно сделать вывод о том, что рабочее место инженера соответствует нормам, ГОСТам и СанПиН. По степени физической тяжести работа инженера относится к категории лёгких работ.

5.2 Производственная безопасность

При работе с электрическими приборами возникают различные вредные факторы, которые негативно воздействуют на организм человека. Также, при работе с электрическими приборами возникают опасные факторы, их перечень представлен в таблице 4.

Таблица 4 – Опасные и вредные факторы при выполнении работ с электрическими приборами

Факторы (ГОСТ)	Этапы работ	Нормативные документы

12.0.003-2015)	Разработка	Изготовление	Эксплуатация	
Отклонение показателей микроклимата в помещении	-	-	+	СП 2.2.4.548-96 [11]
Недостаток естественного освещения	-	-	+	СП 52.13330.2016 [3]
Недостаточная освещенность рабочей зоны	-	-	+	
Повышение уровня электромагнитных излучений	-	-	+	СанПин 2.2.4.3359-16. [4] ГОСТ 12.1.019-2017 [5] ГОСТ 12.1.004-91 ССБТ [6] ГОСТ 12.1.010-76 ССБТ [7]
Повышенная напряженность электромагнитного поля	-	-	+	

5.2.1 Отклонение показателей микроклимата в помещении

Микроклимат является важной характеристикой производственных помещений. В организме человека происходит непрерывное выделение тепла. Одновременно с процессами выделения тепла происходит непрерывная теплоотдача в окружающую среду. Равновесие между выделением тепла и теплоотдачей регулируется процессами терморегуляции, т.е. способностью организма поддерживать постоянство теплообмена с сохранением постоянной температуры тела. Отдача тепла происходит различными видами: излучением, конвекцией, испарением влаги.

Нарушение теплового баланса в условиях высокой температуры может привести к перегреву тела, и как следствие к тепловым ударам с потерей

сознания. В условиях низкой температуры воздуха возможно переохлаждение организма, могут возникнуть простудные болезни, радикулит, бронхит и другие заболевания.

Показателями, которые характеризуют микроклимат рабочей зоны, являются:

- температура воздуха, °С;
- относительная влажность воздуха, %;
- скорость движения воздуха, м/с.

Оптимальные значения этих характеристик зависят от сезона (холодный, тёплый), а также от категории физической тяжести работы. Для математика и программиста она является лёгкой (1а), так как работа проводится сидя, без систематических физических нагрузок.

Оптимальные показатели микроклимата рабочей зоны, согласно СанПиН 2.2.4.548 [2], представлены в таблице 5, допустимые – в таблице 6.

Таблица 5 – Оптимальные показатели микроклимата

Период года	Температура воздуха, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Теплый	23-25	40-60	0,1
Холодный	22-24	40-60	0,1

Таблица 6 – Допустимые показатели микроклимата

Период года	Температура воздуха, °С		Относительная влажность воздуха, %	Скорость движения воздуха, м/с	
	Диапазон ниже оптимальных величин	Диапазон выше оптимальных величин		Диапазон ниже оптимальных величин	Диапазон выше оптимальных величин
Теплый	21,0-22,9	24,1-25,0	15-75	0,1	0,2
Холодный	20,0-21,9	25,1-28,0	15-75	0,1	0,1

Мероприятия по доведению микроклиматических показателей до нормативных значений включаются в комплексные планы предприятий по охране труда. Это такие мероприятия, как:

- механизация и автоматизация производственных процессов, дистанционное управление ими;
- применение технологических процессов и оборудования, исключающих образование вредных веществ или попадания их в рабочую зону;
- установка систем вентиляции, кондиционирования, отопления.

К мероприятиям по оздоровлению воздушной среды в помещении относятся правильная организация вентиляции и кондиционирования воздуха, отопление помещений. Вентиляция может осуществляться естественным и механическим путём. В зимнее время в помещении необходимо предусмотреть систему отопления.

По степени физической тяжести работа программиста относится к категории лёгких работ.

Таким образом, делаем вывод о том, что рабочее место программиста соответствует нормам показателей микроклимата так, как есть соответствие температурным показателям в помещении в холодный период – 23,3 °С, и в теплый период – 24,6 °С. Относительная влажность воздуха составляет 58%, что соответствует диапазону.

5.2.2 Недостаток естественного освещения

Помещения с постоянным пребыванием людей должны иметь естественное освещение.

Без естественного освещения допускается проектировать помещения с временным пребыванием людей, помещения, которые определены соответствующими сводами правил и стандартами организаций на проектирование зданий и сооружений, а также помещения, размещение которых разрешено в подвальных этажах зданий и сооружений.

В учебных и учебно-производственных помещениях общеобразовательных организаций, интернатов, профессиональных образовательных организаций при одностороннем боковом освещении нормируемое значение коэффициента естественной освещенности (КЕО) должно быть обеспечено в расчетной точке, расположенной на пересечении вертикальной плоскости характерного разреза помещения и условной рабочей поверхности на расстоянии 1,2 м от стены, наиболее удаленной от световых проемов и иметь значение не более 0,5%.

Равномерность естественного освещения не нормируется для производственных помещений с боковым освещением.

В аудитории имеется одно окно с одной стороны, чего достаточно для этого помещения, так как в аудитории имеется искусственное освещение.

5.2.3 Повышенная напряженность электромагнитного поля

Измерительные приборы должны подключаться к сети питания имеющей защитное заземление.

Соединять и разъединять вилки, розетки электрических соединений допускается только при выключенном сетевом выключателе.

Основными причинами поражения человека электрическим током могут быть следующие:

- непосредственное прикосновение к токоведущим частям, оказавшимся под напряжением;
- соприкосновение с конструктивными частями, оказавшимися под напряжением.

Электрический ток, проходя через организм человека, оказывает тепловое (ожоги, нагрев сосудов), механическое (разрыв тканей, сосудов при судорожных сокращениях мышц), химическое (электролиз крови), биологическое (раздражение и возбуждение живой ткани) или комбинированное воздействие.

Основными средствами и способами защиты от поражения электрическим током являются:

- недоступность токоведущих частей для случайного прикосновения;
- защитное заземление, зануление или отключение;
- вывешивание предупреждающих надписей;
- контроль за состоянием изоляции электрических установок;
- использование дополнительных средств защиты.

Требования электробезопасности электроустановок производственного и бытового назначения на стадиях проектирования, изготовления, монтажа, наладки, испытаний и эксплуатации, а также технические способы и средства защиты, обеспечивающие электробезопасность электроустановок различного назначения приведены в ГОСТ Р 12.1.019-2017 [5].

5.2.4 Недостаточная освещенность рабочей зоны

Производственное освещение является неотъемлемым элементом условий трудовой деятельности человека. При правильно организованном освещении рабочего места обеспечивается сохранность зрения человека и нормальное состояние его нервной системы, а также безопасность в процессе производства. Производительность труда и качество выпускаемой продукции находятся в прямой зависимости от освещения.

При неудовлетворительном освещении ощущается усталость глаз и переутомление, что приводит к снижению работоспособности. В ряде случаев это может привести к головным болям. Головные боли также могут быть вызваны пульсацией освещения, что в основном является результатом использования электромагнитных пускорегулирующих аппаратов (ПРА) для газоразрядных ламп, работающих на частоте 50 Гц.

Для характеристики естественного освещения используется коэффициент естественной освещенности (КЕО). Величины КЕО для различных помещений лежат в пределах 0,1-12%.

Освещённость на рабочем месте должна соответствовать характеру зрительной работы; равномерное распределение яркости на рабочей поверхности и отсутствие резких теней; величина освещения постоянна во времени (отсутствие пульсации светового потока); оптимальная направленность светового потока и оптимальный спектральный состав; все элементы осветительных установок должны быть долговечны, взрыво-, пожаро-, электробезопасны.

Работа с приборами относится к зрительным работам средней точности для помещений жилых и общественных зданий. Согласно СП 52.13330.2016 [3], такие помещения должны удовлетворять требованиям, указанным в таблице 7.

Таблица 7 - Требования к освещению помещений жилых и общественных зданий при зрительной работе средней точности

Характер зрительной работы	Наименьший или эквивалентный размер объекта различения, мм	Разряд зрительной работы	Подразряд зрительной работы	Относительная продолжительность зрительной работы при направлении зрения на рабочую поверхность, %	Искусственное освещение		Естественное освещение	
					Освещенность на рабочей поверхности от системы общего освещения, ЛК	Коэффициент пульсации освещенности $K_{п, не более}$	КЕО e_n , %, при	
							Верхнем или комбинированном	Боковом
Высокой точности	Более 0,5	В	1	Не менее 70	150	20	2,0	0,5
			2	Менее 70	100	20	2,0	0,5

На рабочем месте соблюдаются необходимые нормы освещенности согласно ГОСТ Р 55710-2013. «Освещение рабочих мест внутри зданий. Нормы и методы измерений». В административных зданиях, а в частности в

помещениях для чтения, письма и обработки данных эксплуатационная освещенность должна составлять 500 ЛК, что соответствует показателям.

5.2.5 Повышенный уровень электромагнитных излучений

В качестве источника электромагнитного излучения в данной работе рассматриваются средства измерения и электрические приборы.

Степень и характер воздействия ЭМП на организм человека зависят: от интенсивности излучения; частоты колебаний; поверхности тела облучаемого; индивидуальных особенностей организма; режима облучения (непрерывный или прерывистый) продолжительности воздействия; комбинированного действия других факторов производственной среды.

В диапазонах промышленной частоты, радиочастот, инфракрасного и частично ультрафиолетового света электромагнитные поля оказывают тепловое воздействие. Перегревание отдельных тканей и органов ведёт к их заболеваниям, а повышение температуры тела на 1 °С и выше вообще не допустимо.

Влияние электромагнитных излучений заключается не только в их тепловом воздействии. Микропроцессы, протекающие в организме под действием излучений, заключаются в поляризации макромолекул тканей и ориентации их параллельно электрическим силовым линиям, что может приводить к изменению свойств молекул; особенно для человеческого организма важна поляризация молекул воды.

Длительное и систематическое воздействие на человека полей ВЧ и УВЧ вызывает:

- повышенную утомляемость;
- головную боль;
- сонливость;
- гипертонию;
- боли в области сердца.

Длительное и систематическое воздействие на человека полей СВЧ вызывает, кроме того:

- изменения в крови;
- катаракту (помутнение хрусталика глаза);
- нервно-психические заболевания [9].

Согласно СанПиН 2.2.4.3359-16 [4], предельно допустимые уровни энергетических экспозиций электромагнитного поля диапазона частот ≥ 30 кГц - 300 ГГц, не должны превышать значений, приведенных в таблице 4.5.

Таблица 4.5 - ПДУ энергетических экспозиций ЭМП диапазона частот ≥ 30 кГц - 300 ГГц

Таблица 8 - Предельно допустимые уровни энергетической экспозиции

Диапазоны частот	Предельно допустимые уровни энергетической экспозиции		
	По электрической составляющей, (В/м) $2 \times \text{ч}$	По магнитной составляющей, (А/м) $2 \times \text{ч}$	По плотности потока энергии (мкВт/см ²) $\times \text{ч}$
30 кГц- 3 МГц	20000,0	200,0	-
3-30 МГц	7000,0	-	-
30-50 МГц	800,0	0,72	-
50-300 МГц	800,0	-	-
300 МГц-300 ГГц	-	-	200,0

Для понижения уровня напряженности электромагнитного поля следует использовать мониторы с пониженным уровнем излучения, устанавливать защитные экраны, устранять неисправности. А для профилактики компьютерного зрительного синдрома, улучшения визуальных показателей видеомониторов, повышения работоспособности, снижения зрительного утомления возможно применение защитных очков со спектральными фильтрами, разрешенных Минздравом России для работы с персональным компьютером.

На рабочем месте программиста используются мониторы с пониженным уровнем излучения. Таким образом можно сделать вывод о том, что рабочее место соответствует нормам СанПиН 2.2.4.3359-16 [4].

5.3 Экологическая безопасность

В настоящее время, когда встает проблема рационального использования природных ресурсов, охраны окружающей среды, уделяется большое внимание организации разумного воздействия на природу. Необходимо совершенствовать технологические процессы с целью сохранения окружающей среды от вредных выбросов.

В связи с тем, что основным средством работы являются средства измерения и электрические приборы, серьезной проблемой является электропотребление. Это влечет за собой общий рост объема потребляемой электроэнергии. Для удовлетворения потребности в электроэнергии, приходится увеличивать мощность и количество электростанций. Соответственно, рост энергопотребления приводит к таким экологическим нарушениям, как глобальное потепление климата, загрязнение атмосферы и водного бассейна Земли вредными и ядовитыми веществами, опасность аварий в ядерных реакторах, изменение ландшафта Земли. Целесообразным является разработка и внедрение систем с малым потреблением энергии.

В аудитории не ведется никакого производства. К отходам, производимым в помещении можно отнести бытовой мусор.

Сточные воды здания относятся к бытовым сточным водам. За их очистку отвечает городской водоканал.

Основной вид мусора – это отходы печати, бытовой мусор (в т. ч. люминесцентные лампы), неисправное электрооборудование, коробки от техники, использованная бумага. Утилизация отходов печати вместе с бытовым мусором происходит в обычном порядке.

Утилизация средств измерений и электрических приборов согласно ФЗ от 24.06.1998 №89-ФЗ (ред. от 25.12.2018) «Об отходах производства и потребления» [12] и распоряжению Правительства РФ от 25.07.2017 №1589-р [13] осуществляется сотрудниками лаборатории и предусматривает следующие этапы:

1. Правильное заполнение акта списания с указанием факта невозможности дальнейшей эксплуатации перечисленной в акте измерительной техники, о чем имеется акт технического осмотра;

2. Осуществление списания перечисленной в акте измерительной техники с баланса предприятия с указанием в бухгалтерском отчете, так как утилизация возможна для осуществления только после окончательного списания;

3. Непосредственно утилизация измерительной техники с полным демонтажем устройств на составляющие детали с последующей сортировкой по видам материалов и их дальнейшей передачей на перерабатывающие заводы.

5.4 Безопасность в чрезвычайных ситуациях

Чрезвычайными ситуациями в подобных помещениях могут быть пожары. Основы пожарной безопасности определены по ГОСТ 12.1.004-91 [6] и ГОСТ 12.1.010-76 [7].

Все производства по пожарной опасности подразделяются на 5 категорий: А, Б, В, Г, Д. Аудитория, в которой будет выполняться работа, относится к категории В.

Причинами пожара могут быть:

- токи короткого замыкания;
- электрические перегрузки;
- выделение тепла, искрение в местах плохих контактов при соединении проводов;
- курение в неположенных местах.

Тушение горящего электрооборудования под напряжением должно осуществляться имеющимися огнетушителями ОУ-5. Чтобы предотвратить пожар в аудитории, необходимо:

- содержать помещение в чистоте, убирать своевременно мусор. По окончании работы поводится влажная уборка всех помещений;

- работа должна проводиться только при исправном электрооборудовании;
- на видном месте должен быть вывешен план эвакуации из помещения с указанием оборудования, которое нужно эвакуировать в первую очередь;
- уходящий из помещения последним должен проверить выключены ли нагревательные приборы, электроприборы и т.д и отключение силовой и осветительной электрической сети.

Также необходимо соблюдение организационных мероприятий:

- правильная эксплуатация приборов, установок;
- правильное содержание помещения;
- противопожарный инструктаж сотрудников аудитории;
- издание приказов по вопросам усиления ПБ;
- организация добровольных пожарных дружин, пожарно-технических комиссий;
- наличие наглядных пособий и т.п.

В случаях, когда не удастся ликвидировать пожар самостоятельно, необходимо эвакуироваться вслед за сотрудниками по плану эвакуации и ждать приезда специалистов, пожарников. При возникновении пожара должна сработать система пожаротушения, издав предупредительные сигналы, и передав на пункт пожарной станции сигнал о ЧС, в случае если система не сработала, по каким-либо причинам, необходимо самостоятельно произвести вызов пожарной службы по телефону 101, сообщить место возникновения ЧС и ожидать приезда специалистов.

5.5 Выводы по разделу «Социальная ответственность»

Таким образом можно сделать вывод по всему разделу «Социальная ответственность», что на рабочем месте показатели микроклимата, освещенности и электромагнитных излучений находятся в норме, что говорит о безопасности для человека. Анализ выявленных опасных факторов рабочего

помещения показал, что электробезопасность, пожаробезопасность и экологическая безопасность находятся под контролем, и все необходимые меры для обеспечения безопасности принимаются.

6. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

6.1 Предпроектный анализ

6.1.1 Потенциальные потребители результатов исследований

Любой ученый заинтересован в тех потребителях, которые приобретут его разработку по цене, обеспечивающей получение прибыли в достаточном объеме и на протяжении длительного периода. Поэтому необходимо четко представлять, кто является конечным потребителем разработки или кто может стать им в будущем и по какой причине они эту разработку приобретают, т. е. определить свой целевой рынок. [1]

Разработка, проведенная в ходе данной работы, позволяет автоматизировать множество процессов для сотрудников предприятий. Это позволяет экономить время на обучение персонала, время выполнения рутинных задач сотрудниками.

В нашем случае сегментами рынка выступают размер компании и требуемый для нее функционал.

Таблица 9 - Потенциальные потребители результатов исследования

		По требуемому функционалу		
		Бот только для поиска информации	Бот с расширенными возможностями (отправка сообщений, заполнение документов и пр.)	Бот с расширенным функционалом и гибкой настройкой
Размер компании	Крупные			ПАО «НК «Роснефть»
	Средние		ОАО «Томскнефть»	
	Мелкие	ООО «Стрежевской		

		нефтеперерабаты вающий завод»		
--	--	----------------------------------	--	--

6.1.2 Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения

Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения позволяет провести оценку сравнительной эффективности научной разработки и определить направления для ее будущего повышения. Для этого составлена оценочная карта, приведенная в таблице 2. Конкурент 1 - компания “Singularika”, конкурент 2- компания “Jetstyle” [2].

Таблица 10 – Оценочная карта для сравнения конкурентных технических разработок

Критерии оценки	Вес критерия	Баллы			Конкурентно-способность		
		Бф	Бк1	Бк2	Кф	Кк1	Кк2
Технические критерии оценки ресурсоэффективности							
1 Способствует увеличению скорости выполнения рутинных задач	0,25	5	4	4	1,25	1	1
2 Удобство в эксплуатации чат-бота (соответствует требованиям потребителей все функции в одном приложении)	0,20	5	5	5	1	1	1
3 Простота эксплуатации чат-бота	0,17	5	5	3	0,85	0,85	0,51

(понятный интерфейс, понятные команды, одно приложение)							
Экономические критерии оценки эффективности							
1 Цена	0,13	5	4	3	0,65	0,52	0,39
2 Предполагаемый срок эксплуатации	0,1	5	4	3	0,5	0,4	0,3
3 Конкурентоспособность продукта	0,09	5	4	4	0,45	0,36	0,36
4 Уровень проникновения на рынок	0,06	4	5	5	0,24	0,3	0,3
Итого	1	34	32	27	4,94	4,43	3,86

Анализ конкурентных технических решений определяется по формуле:

$$K = \sum V_i \cdot B_i$$

где K – конкурентоспособность научной разработки или конкурента; V_i – вес показателя (в долях единицы); B_i – балл i -го показателя. Таким образом, конкурентоспособность разработки составила 4,94, в то время как двух других аналогов 4,43 и 3,86 соответственно. Результаты показывают, что данная научно-исследовательская разработка является конкурентоспособной и имеет преимущества по большинству критериям [33].

6.1.3 SWOT-анализ

SWOT – Strengths (сильные стороны), Weaknesses (слабые стороны), Opportunities (возможности) и Threats (угрозы) – представляет собой комплексный анализ научно-исследовательского проекта. SWOT-анализ применяют для исследования внешней и внутренней среды проекта.

Результаты первого этапа SWOT-анализа представлены в таблице 3.

Таблица 11 – Оценочная карта для сравнения конкурентных технических разработок

	<p>Сильные стороны:</p> <p>С1: Наличие опытного научного руководителя и консультанта в продвижении продукта</p> <p>С2: Построение продукта на бесплатных платформах, с открытым исходным кодом</p> <p>С3: Большое количество функций в продукте</p>	<p>Слабые стороны:</p> <p>Сл1: Малый опыт разработчика</p> <p>Сл2: Трудоемкость процесса разработки и тестирования продукта</p> <p>Сл3: Малое количество потребителей</p>
<p>Возможности:</p> <p>В1: Актуальность чат-ботов</p> <p>В2: Ориентация именно на крупные предприятия, с подготовленным функционалом, в отличии от конкурентов</p> <p>В3: Легкость внедрения продукта</p>	<p>С1В1В2 – позволит оптимизировать продвижение продукта.</p> <p>С2С3В3 – позволяет удовлетворить потребности крупных предприятий одним приложением.</p>	<p>Так как продукт должен легко внедряться в готовые корпоративные системы, необходимо предусмотреть и подготовить различные версии продукта, а так же полностью протестировать</p>
<p>Угрозы:</p> <p>У1: Отсутствие спроса на конечный продукт.</p> <p>У2: Развитая конкуренция в области разработки чат-ботов.</p> <p>У3: Наличие альтернативных подходов.</p>	<p>Большое количество функций в продукте, а так же его безопасность позволит заинтересовать потребителей, а так же победить конкурентов.</p>	<p>Специализация именно на крупные предприятия позволяет разработать готовую платформу, и в дальнейшем внедрять в кратчайшие сроки, что не могут сделать конкуренты, так как процесс разработки</p>

		долгий.
--	--	---------

В ходе анализа были изучены сильные стороны и возможности проекта, а также обозначены его слабые стороны и угрозы, которые необходимо учесть при реализации проекта

6.1.4 Оценка готовности проекта к коммерциализации

На какой бы стадии жизненного цикла не находилась научная разработка полезно оценить степень ее готовности к коммерциализации и выяснить уровень собственных знаний для ее проведения (или завершения). На таблице 3 показано оценки степени готовности научного проекта к коммерциализации [3].

Оценки степени проработанности научного проекта трактуется следующим образом:

- Не проработано;
- Проработано слабо;
- Выполнено, но качество под сомнением;
- Выполнено качественно;
- Имеется положительного заключение независимого эксперта.

Оценка уровня имеющихся знаний у разработчика определяется в соответствии со следующей системой баллов:

- Не знаком или знаком мало;
- Знаком с теорией;
- Знаком с теорией и практическими примерами применения;
- Знаком с теорией и самостоятельно выполняет;
- Знаком с теорией, выполняет, может консультировать.

Таблица 12 - Оценки степени готовности научного проекта к коммерциализации

Наименование	Степень проработанности	Уровень имеющихся знаний у разработчика

	научного проекта	
Определен имеющийся научно-технический задел	4	4
Определены перспективные направления коммерциализации научно-технического задела	5	5
Определены отрасли и технологии (товары, услуги) для предложения на рынке	5	5
Определена товарная форма научно-технического задела для представления на рынок	2	2
Определены авторы и осуществлена охрана их прав	5	5
Проведена оценка стоимости интеллектуальной собственности	5	5
Проведены маркетинговые исследования рынка сбыта	4	4

Разработан бизнес-план коммерциализации научной разработки	1	1
Определены пути продвижения научной разработки на рынок	4	4
Разработана стратегия (форма) реализации научной разработки	3	4
Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	2	2
Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	4	4
Проработаны вопросы финансирования коммерциализации научной разработки	4	4
Имеется команда для коммерциализации научной разработки	4	4
Проработан механизм реализации научного проекта	4	4

ИТОГО БАЛЛОВ	56	57
---------------------	----	----

Оценка готовности научного проекта к коммерциализации (или уровень имеющихся знаний у разработчика) определяется по формуле:

$$B_{\text{сум}} = \sum B_i,$$

где $B_{\text{сум}}$ – суммарное количество баллов по каждому направлению;

B_i – балл по i -му показателю.

Значение $B_{\text{сум}}$ **позволяет** говорить о мере готовности научной разработки и ее разработчика к коммерциализации. Перспективность разработки выше среднего, так как значение и знания разработчика достаточны для успешной ее коммерциализации [5].

6.1.5 Методы коммерциализации результатов научно-технического исследования

При коммерциализации научно-технических разработок продавец (а это, как правило, владелец соответствующих объектов интеллектуальной собственности), преследует вполне определенную цель, которая во многом зависит от того, куда в последующем он намерен направить (использовать, вложить) полученный коммерческий эффект. Это может быть получение средств для продолжения своих научных исследований и разработок (получение финансирования, оборудования, уникальных материалов, других научно-технических разработок и т. д.), одноразовое получение финансовых ресурсов для каких-либо целей или для накопления, обеспечение постоянного притока финансовых средств, а также их различные сочетания. При этом время продвижения товара на рынок во многом зависит от правильности выбора метода коммерциализации. Для разработанного проекта инжиниринг является самым оптимальным методом коммерциализации. Инжиниринг – самостоятельный вид коммерческих операций предполагает предоставление на основе договора инжиниринга одной стороной, именуемой консультантом, другой стороне, именуемой заказчиком, комплекса или отдельных видов

инженерно-технических услуг, связанных с проектированием, строительством и вводом объекта в эксплуатацию, с разработкой новых технологических процессов на предприятии заказчика, усовершенствованием имеющихся производственных процессов вплоть до внедрения изделия в производство и даже сбыта продукции [1].

6.2 Инициализация проекта

6.2.1 Цели и результат проекта

Инициация проекта определяет изначальные цели проекта или содержание работы. Также, в данном разделе определяются изначальные экономические – финансовые затраты. По данной работе также надо определить внутренние и внешние заинтересованные стороны проекта, которые влияют на общий результат научного проекта [5].

Таблица – 13 – Заинтересованные стороны проекта

Заинтересованные стороны проекта	Ожидания заинтересованных сторон
Крупные предприятия, обладающие сложными корпоративными сетями, с большим количеством данных	Внедрение продукта за короткий промежуток времени, наличие функций автоматизирующих рутинные задачи.

Таблица 14 – Цели и результат проекта

Цели проекта:	Целью работы является разработка встраиваемого чат-бота в готовые корпоративные системы крупных предприятий
Ожидаемые результаты проекта:	Разработанное ПО, готовое к внедрению в корпоративные системы
Критерии приемки результата проекта:	- Качество обработки запросов

	пользователем; - Универсальность решения.
Требования к результату проекта:	- Отсутствие угроз утечки корпоративной информации; - Скорость внедрения готового продукта в корпоративные системы; - Надежность системы (отсутствие влияния внешней среды на приложение, отказоустойчивость). - Возможность общения с ботом на естественном языке

6.2.2 Организационная структура проекта

На данном этапе работы необходимо решить следующие вопросы: кто будет входить в рабочую группу данного проекта, определить роль каждого участника в данном проекте, а также прописать функции, выполняемые каждым из участников и их трудозатраты в проекте [33].

Таблица 15 – Рабочая группа проекта

№ п/п	Фамилия И.О. место работы, должность	Роль в проекте	Функции в проекте	Трудозатра ты, час.
1	Семенов М.Е.	Руководите ль проекта	Координир ует деятельность участников проекта	28
2	Чурсин Г.С.	Исполните ль проекта	Выполнени е разработки	546

6.3 Планирование управления научно-техническим проектом

Группа процессов планирования состоит из процессов, осуществляемых для определения общего содержания работ, уточнения целей и разработки последовательности действий, требуемых для достижения данных целей [5].

6.3.1 План проекта

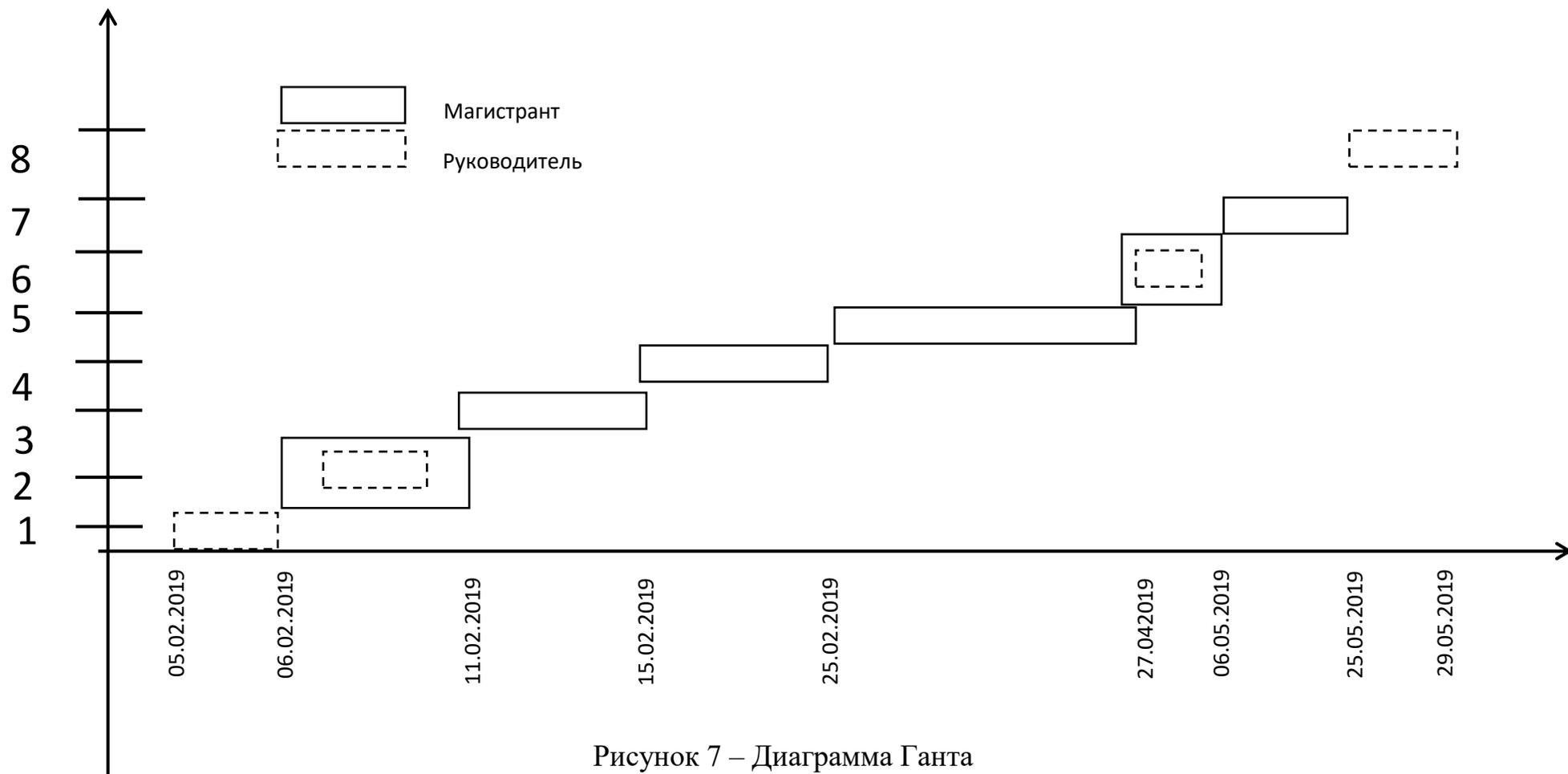
Диаграмма Ганта – горизонтальный ленточный график, на котором работы по теме представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ.

График строится с разбивкой по месяцам и декадам (10 дней) за период времени дипломирования. При этом работы на графике выделяются различной штриховкой в зависимости от исполнителей, ответственных за ту или иную работу [5].

Таблица 16 – Календарный график проведения НИР по теме

Название	Время, дни	Дата начала работ	Дата окончания работ	Состав участников
Составление и утверждение технического задания	2	05.02.19	06.02.19	Руководитель
Подбор и изучение материалов по теме	4	07.02.19	11.02.19	Руководитель / Инженер
Описание объекта исследования	4	12.02.19	15.02.19	Инженер

Сбор и анализ исходных данных исследуемого объекта	6	16.02.19	23.02.19	Инженер
Разработка прототипа программы	53	25.02.19	27.04.19	Инженер
Тестирование	5	29.05.19	06.05.19	Руководитель / инженер
Оформление пояснительной записки	19	06.05.19	25.05.19	Инженер
Оценка эффективности проделанной работы и проверка пояснительной работы	3	27.05.19	29.05.19	Руководитель
Итого	Магистрант			91
	Руководитель			14



6.3.2 Бюджет научно-технического исследования (НТИ)

При планировании бюджета научных исследований должны быть обеспечены полным и точным отражением всех запланированных затрат, необходимых для ее реализации

6.3.4.1 Затраты на оплату труда работников, непосредственно занятых выполнением НИОКР

Статья включает основную заработную плату работников, непосредственно занятых выполнением проекта, (включая премии, доплаты) и дополнительную заработную плату.

$$C_{зп} = З_{осн} + З_{доп},$$

где $З_{осн}$ – основная заработная плата;

$З_{доп}$ – дополнительная заработная плата.

Основная заработная плата ($З_{осн}$) руководителя от университета:

$$З_{осн} = З_{дн} * T_{раб}$$

где $З_{осн}$ – основная заработная плата работника; T_p – продолжительность работ, выполняемых научно-техническим работником, раб. дн.; $З_{дн}$ – среднедневная заработная плата работника, руб.

Среднедневная заработная плата рассчитывается по формуле:

$$З_{дн} = \frac{З_m * M}{F_d}$$

где $З_m$ – месячный должностной оклад работника, руб.; M – количество месяцев работы без отпуска в течение года: при отпуске в 48 раб.дня за полгода $M = 10,4$ месяца, 6-дневная неделя; F_d – действительный годовой фонд рабочего времени научно-технического персонала, раб. дн.

Таблица 17 – Баланс рабочего времени

Показатели рабочего времени	Руководитель	Магистрант

Календарное число дней	365	365
Количество нерабочих дней	52	52
- выходные дни	14	14
- праздничные дни		
Потери рабочего времени	48	48
- отпуск		
Действительный годовой фонд рабочего	251	251

Месячный должностной оклад работника:

$$Z_m = Z_{TC} * (1 + k_{пр} + k_d) * k_p,$$

где Z_{TC} – заработная плата по тарифной ставке, руб.;

$k_{пр}$ – премиальный коэффициент, равный 0,3 (т.е. 30% от Z_{TC});

k_d – коэффициент доплат и надбавок составляет примерно 0,2 – 0,5 (в НИИ и на промышленных предприятиях – за расширение сфер обслуживания, за профессиональное мастерство, за вредные условия: 15-20 % от Z_{TC});

k_p – районный коэффициент, равный 1,3 (для Томска).

Месячный должностной оклад работника:

$$Z_m = Z_б * k_p,$$

где $Z_б$ – базовый оклад, руб.; k_p – районный коэффициент, равный 1,3 (для Томска).

Таблица 18 – Расчёт основной заработной платы

	$Z_б$, руб.	k_p	Z_m , руб.	$Z_{дн}$, руб.	T_p раб. дн.	$Z_{осн}$, руб.
Руководитель	33664	1,3	43763,2	1813,3	4,67	8468,5
Студент	12663	1,3	16461,9	682,1	91	62071,1

6.3.4.2 Дополнительная заработная плата

Дополнительная заработная плата включает оплату за непроработанное время (очередной и учебный отпуск, выполнение государственных обязанностей, выплата вознаграждений за выслугу лет и т.п.) и рассчитывается исходя из 10-15% от основной заработной платы, работников, непосредственно участвующих в выполнении темы:

$$Z_{\text{доп}} = K_{\text{доп}} * Z_{\text{осн}},$$

где $Z_{\text{доп}}$ – дополнительная заработная плата, руб.;

$K_{\text{доп}}$ – коэффициент дополнительной зарплаты;

$Z_{\text{осн}}$ – основная заработная плата, руб.

Таблица 19 – Заработная плата исполнителей НИОКР

Заработная плата	Руководитель	Инженер
Основная заработная плата	8468,5	62071,1
Дополнительная заработная плата (10%)	846,85	6207,11
Итого по статье $C_{\text{зп}}$	9315,35	68278,21

6.3.4.4 Отчисления во внебюджетные фонды (страховые отчисления)

В данной статье расходов отражаются обязательные отчисления по установленным законодательством Российской Федерации нормам органам государственного социального страхования (ФСС), пенсионного фонда (ПФ) и медицинского страхования (ФФОМС) от затрат на оплату труда работников.

Величина отчислений во внебюджетные фонды определяется исходя из следующей формулы:

$$Z_{\text{внеб}} = k_{\text{внеб}} * Z_{\text{осн}},$$

где $k_{\text{внеб}}$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

На 2015 г. в соответствии с Федеральным законом от 24.07.2009 №212-ФЗ установлен размер страховых взносов равный 30,2%. На основании пункта 1 ст.58 закона №212-ФЗ для учреждений, осуществляющих образовательную и научную деятельность в 2015 году водится пониженная ставка – 27,1%.

Таблица 20 – Отчисления во внебюджетные фонды

	Руководитель	Магистрант
Зарплата	9315,35	68278,21
Отчисления во внебюджетные фонды	2524,46	18503,4

6.3.4.4 Прочие прямые затраты

Компьютер потребляет 0,15 кВт*ч. В Томске 1 кВт*ч стоит 5,8 рублей.

На работу за компьютером было затрачено 574 часа. Значит стоимость потраченной электроэнергии $0,15 \cdot 574 \cdot 5,8 = 499,38$ руб.

6.3.4.4 Накладные расходы

В данную статью входят расходы на содержание аппарата управления и общехозяйственных (общеуниверситетских) служб, которые в равной степени относятся ко всем выполняемым НИОКР. По этой статье учитываются оплата труда административно-управленческого персонала, содержание зданий, оргтехники и хоз. инвентаря, амортизация имущества, расходы по охране труда и подготовке кадров [5].

Накладные расходы лаборатории 2 корпуса НИ ТПУ составляют 80-90% от суммы основной и дополнительной заработной платы, работников, непосредственно участвующих в выполнении темы. Расчет накладных расходов ведется по следующей формуле:

$$C_{\text{накл}} = K_{\text{накл}} \cdot (Z_{\text{осн}} + Z_{\text{доп}}),$$

где $K_{\text{накл}}$ – коэффициент накладных расходов. $C_{\text{накл}} = 80\% \cdot C_{\text{зп}} = 62031,65$

На основании полученных данных по отдельным статьям затрат составляется калькуляция плановой себестоимости НИР по форме, приведенной в таблице 21.

Таблица 21 – Калькуляции плановой себестоимости НИР

Наименование статей затрат	Сумма, руб
1. Затраты на оплату труда работников, непосредственно занятых созданием НИОКР	77593,56
2. Отчисления во внебюджетные фонды	21027,86
3. Прочие прямые затраты	499,38
4. Накладные расходы	62031,65
Итого себестоимость НИР	161152,45

6.4 Реестр рисков проекта

Идентифицированные риски проекта включают в себя возможные неопределенные события, которые могут возникнуть в проекте и вызвать последствия, которые повлекут за собой нежелательные эффекты.

Таблица 22 – Риски

№	Риск	Потенциальное воздействие	Вероятность наступления (1-5)	Влияние риска (1-5)	Уровень риска *	Способы смягчения риска	Условия наступления
1	Отсутствие спроса со стороны ключевых	Отсутствии возможно сти	2	4	Средний	- Разработка политики по	Отсутствие заинтересованности в конкретном

	потребител ей для приобретен ия продукта	продаж продукци и				продвижен ию продукции на рынке - подготовка к смене ориентаци и продукта	продукте
2	Технологи ческие риски, связанные с использова нием ПО, сторонних разработчи ков	Не завершен ность продукта	1	2	Низк ий	Рассмотре ние аналогичн ых ПО используе мых в разработке	Изменение политики распростране ния используемо го ПО
3	Сбои в работке готовой продукции	Проблем ы с внедрение программ ы в готовые корпорати вные системы	2	2	Низк ий	Проводить как полное так модульное тестирован ие	Отсутствие полноценног о тестирования
4	Конкурент	Снижение	2	4	Сред	-	- Высокая

	ия на рынке	спроса на разработа нное ПО			ний	Подготовк а к снижению стоимости продукции - Поиск дополните льных клиентов из других категорий	цена продукта - Отсутствие необходимог о функционала
--	----------------	-----------------------------------	--	--	-----	---	---

6.5 Оценка сравнительной эффективности исследования

Определение эффективности происходит на основе расчета интегрального показателя эффективности научного исследования.

Таблица 23 - Сравнительная оценка характеристик вариантов исполнения проекта

	Весовой коэффициент параметра	Текущий проект	Аналог 1	Аналог 2
1 Способствует росту производительности труда	0,1	5	4	4
2 Удобство в эксплуатации (соответствует требованиям)	0,2	5	5	5

потребителей)				
3 Простота эксплуатации	0,15	5	5	5
4 Безопасное хранение данных	0,25	5	4	4
5 Скорость внедрения	0,3	5	3	3
ИТОГО	1	25	21	21

$$I_{\text{тп}} = 5*0,1+5*0,2+5*0,15+5*0,25+5*0,3 = 5$$

$$\text{Аналог 1} = 4*0,1+5*0,2+5*0,15+4*0,25+3*0,3 = 4,05$$

$$\text{Аналог 2} = 4*0,1+5*0,2+5*0,15+4*0,25+3*0,3 = 4,05$$

Из приведенных расчетов выявлено, что текущий проект по интегральному показателю ресурсоэффективности является выгодным и превосходит аналоги. Так как данный проект является только научной разработкой и началом исследования, то интегральный финансовый показатель разработки рассчитать не представляется возможным.

Выводы по разделу «Финансовый менеджмент»

В случае внедрения программного комплекса происходит снижение времени на выполнение рутинных задач персонала, так как программный комплект позволяет обходить использование корпоративных систем, и выполнять необходимые задачи как с компьютера, так и с телефона. Кроме того, снижается время на обучения персонала по использованию корпоративных систем. Эффект от экономии времени, приводит к снижению затрат на обучение персонала.

Заключение

1. Проведен анализ существующих технологий и решений.
2. Составлен алгоритм обработки естественного языка.
3. Разработан прототип чат-бота, способный принимать запросы на естественном языке. Добавлена поддержка необходимого функционала для автоматизации офисных задач. На данный момент, первостепенной проблемой является долгая обработка запросов с опечатками. Поиск вариантов замены требует порядка 3 секунд. Возможные решения: уменьшение словаря, нормализация слов перед проверкой на существование, использование более низкоуровневых языков для работы с файлами, смена персонального компьютера на более мощный.
4. Добавлена функция корректировки опечаток, вследствие которых бот выводит данные после исправления ошибок.
5. Проведено тестирование отдельных компонент. Для тестирования использовался мессенджер Telegram, так как не требует глубоких настроек и установки серверов. Выявлено корректное поведение программы на разных типах запросов.

Список публикаций

1. Чурсин Г.С., Исмоилов Н.И., Доржиев А.В. Встраиваемый чат-бот в готовые корпоративные системы // Перспективы развития фундаментальных наук: сборник трудов XVI Международной конференции студентов, аспирантов и молодых ученых. - Томск, 2019.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Б. О. Близнюк, Л. В. Васильева, И. Д. Стрельников, Д. С. Ткачук Современные методы обработки естественного языка // Вестник Харьковского национального университета имени В.Н. Каразина. – 2017. – С.14-26.
2. Ю. В. Давыдова Алгоритм нечеткого текстового поиска в виртуальных социальных сетях // International Journal of Open Information Technologies ISSN. - Москва: Лаборатория Открытых Информационных Технологий факультета ВМК МГУ им. М.В. Ломоносова, 2018.
3. Lucian Sasu A Probabilistic Model For Spelling Correction // Bulletin of the Transilvania University of Brasov. - 2011. - Series III: Mathematics, Informatics, Physics.
4. Mislēvičs, A., Grundspenķis, J., Rollande, R. A Systematic Approach to Implementing Chatbots in Organizations – RTU Leo Showcase. In: Joint Proceedings of the BIR 2018 Short Papers, Workshops and Doctoral Consortium co-located with 17th International Conference Perspectives in Business Informatics Research (BIR 2018). CEUR Workshop Proceedings. Vol. 2218, Sweden, Stockholm, 24-26 September, 2018. Aachen: RWTH, 2018, pp.356-365. ISSN 1613-0073.
5. Chaitrali S. Kulkarni, Amruta U. Bhavsar, Savita R. Pingale Bank Chat Bot – An Intelligent Assistant System Using NLP and Machine Learning // International Research Journal of Engineering and Technology. - 2017. - №05.
6. Atish Pawar, Vijay Mago Calculating the similarity between words and sentences using a lexical database and corpus statistics // IEEE Transactions On Knowledge And Data Engineering . - 2018.
7. Roshan Khan Standardized Architecture for Conversational Agents a.k.a. ChatBots // International Journal of Computer Trends and Technology (IJCTT). - 2017. - №50.
8. Scott A. Crossley, Laura K. Allen, Kristopher Kyle & Danielle S. McNamara (2014) Analyzing Discourse Processing Using a Simple Natural Language Processing Tool, Discourse Processes, 51:5-6, 511-534

9. Martin Dostal, Karel Jeřek Automatic Keyphrase Extraction based on NLP and Statistical Methods // Annual International Workshop on Databases, Texts, Specifications and Objects. - Czech Republic: DATESO, 2011.
10. Pandiselvam. P, Marimuthu. T, Lawrance Raj A Comparative Study on String Matching Algorithm of Biological Sequences // International Conference on Intelligent Computing. 2014.
11. Мокроусов М.Н. Разработка и исследование методов и системы семантического анализа естественно-языковых текстов: дис. ... канд. тех. наук: 05.13.01. - Ижевск, 2010.
12. Давыдова Ю.В. Модель ошибок для нечеткого текстового поиска в задаче мониторинга виртуальных социальных сетей для обеспечения информационно-психологической безопасности личности // Современные информационные технологии и ИТ-образование . - 2017. - №3.
13. Najlah Gali, Radu Marinescu-Istodor, Damien Hostettler Framework for syntactic string similarity measures // Expert Systems with Applications. - Louisiana: 2019.
14. Menal Dahiya A Tool of Conversation: Chatbot // International journal of computer sciences and engineering. - 2017. - №5.
15. Gilberto Ayala-Bastidas, Leonardo Garrido A Knowledge-based Methodology for Building a Conversational Chatbot as an Intelligent Tutor // Mexican International Conference on Artificial Intelligence. - Mexico City: 2018.
16. Ванюшкин А.С., Гращенко Л.А. Методы и алгоритмы извлечения ключевых слов // Новые информационные технологии в автоматизированных системах. - 2016.
17. Шереметьева, С.О. Методы и модели автоматического извлечения ключевых слов / С.О. Шереметьева, П.Г. Осминин // Вестник Южно-Уральского государственного университета. - 2015. - Т. 12, № 1. - С. 76–81.
18. Бойцов Л.М. Классификация и экспериментальное исследование современных алгоритмов нечеткого словарного поиска [текст] // Электронные

библиотеки: перспективные методы и технологии, электронные коллекции:
RCDL2004: тр. 6-й Всеросс. науч. конф.

19. Sameera A. Abdul-Kader, Dr. John Woods Survey on Chatbot Design Techniques in Speech Conversation Systems // (IJACSA) International Journal of Advanced Computer Science and Applications. - 2015. - №6.

20. Steven Bird, Ewan Klein, Edward Loper Natural Language Processing with Python. - Sebastopol: O'RIELLY, 2009.

21. ГОСТ 12.2.032 -78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования. М.: Изд-во стандартов, 1986. – 9 с.

22. Санитарно - эпидемиологические правила и нормативы: СанПиН 2.2.4.548–96 Гигиенические требования к микроклимату производственных помещений. М.: Минздрав России, 1997. – 20 с.

23. Свод правил: СП 52.13330.2016. Естественное и искусственное освещение. М.: Минрегион России, 2011. – 74 с.

24. Санитарно - эпидемиологические правила и нормативы: СанПиН 2.2.4.3359-16. Санитарно-эпидемиологические требования к физическим факторам на рабочих местах. М.: Минздрав России, 2003. – 39 с.

25. ГОСТ 12.1.019-2017 ССБТ Электробезопасность. Общие требования и номенклатура видов защиты. – М.: Изд-во стандартов, 2010. – 28 с.

26. ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования. – М.: Изд-во стандартов, 2006. – 67 с.

27. ГОСТ 12.1.010-76 ССБТ. Взрывобезопасность. Общие требования. – М.: Изд-во стандартов, 2003. – 6 с.

28. Трудовой кодекс Российской Федерации от 30.12.2001 N 197- ФЗ. Официальный текст. - М. : Пропаганда : Омега- Л, 2002. - 176 с. : ил. - (Российская правовая библиотека).

29. В.С. Алексеев, И.В. Ткаченко, О.И. Жидкова. Безопасность жизнедеятельности. Сборник лекций. – М.: Эксмо, 2008. – 160 с.

30. Санитарно – эпидемиологические правила и нормативы: СанПиН 2.2.2/2.4.1340-03 Гигиенические требования к персональным электронно-вычислительным машинам и организации работы. М.: Минздрав России, 2003. – 56 с.
31. Санитарные правила и нормы СанПиН 2.2.4.548-96 "Гигиенические требования к микроклимату производственных помещений".
32. ФЗ от 24.06.1998 №89-ФЗ (ред. от 25.12.2018) «Об отходах производства и потребления».
33. Распоряжение Правительства РФ от 25.07.2017 №1589-р.
34. Гаврикова Н.А. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение: учебно-методическое пособие. – Томск: Изд-во Томского политехнического университета, 2014. – 73 с. 99 13.
35. Рыжакина Т.Г. Бизнес-планирование: Методические указания по выполнению курсовой работы по курсу «Управление и организация производством». – Томск: Изд-во Томского политехнического университета, 2013. – 38 с.

Приложение А

Chapter 1

Review of the literature on the use of chatbots in corporate systems.

Студент:

Группа	ФИО	Подпись	Дата
0ВМ71	Чурсин Геннадий Сергеевич		

Консультант – лингвист отделения ШБИП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель ОИЯ	Кабрышева Оксана Павловна			

Literature review

In article [1] the main problems of natural language processing are considered. The main directions of processing, methods, tools and the libraries available to time are analyzed at the moment. Two experiments where these techniques were applied to the solution of real tasks are made - analysis of tonality of a news background of some cryptocurrencies and the analysis of factors about cooperation of the companies. In conclusion, the analysis of text data has great practical value in the modern world.

In work [2] the algorithm consists of two stages – scanning and verification is presented. Filtration of the text is carried out at a stage of scanning, for the purpose of an exception of consideration of the messages which definitely are not containing required words. Other messages are checked at a verification stage. Integration into search of linguistic rules and statistics of mistakes allows to keep sufficient accuracy at the same time. Testing was held on selection of texts of messages of the General Internet case of Russian.

Article [3] describes processing of mistakes and replacement of words. Word corrections are widely encountered in various applications (google, word, etc.). At the same time they provide the most probable replacements for words which, obviously are wrong. The described algorithm is checked on a mere verbiage from news articles of the New York Times. As a result, the author draws a conclusion that the theoretical component of a method and results of tests show that the algorithm is suitable for spelling.

Chatbot as a means of employee work automation

The chatbot is the program. The aim of the program is communication with the person in the language clear to him. Both the natural language, and the prepared teams and templates can be understood as a language.

The concept “chatbot” was used by Michael Molding in 1994 for the first time. He created the program called Julia which supported communication in the natural language. The program could also self-taught except communication.

However, the first chatbots prototypes were developed long before that. In 1964, MIT professor Joseph Weinbaum created the ELIZA program, which became the first chat-program. The task of the program was to imitate the user's communication with the psychotherapist. The Parry program was invented later which chatbot imitated the speech of a mentally ill person and presented himself as a patient. 48% of professional psychiatrists could not guess that they communicated with the machine.

In the late 90s and early 2000s, chatbots began to spread on the IM and IRC (Internet Relay Chat) platforms. They were used to moderate the chat or conversation.

More advanced chatbots appeared later. In 2006, the Watson program appeared. A feature of the program was the ability to recognize not only text, but also voice messages. This technology served as the basis for creating many voice assistants.

Since 2010 large companies have begun to use and implement their chatbots: Siri, Google Now. Also Amazon and Microsoft presented their developments.

The main task of chatbots is to fulfill user requests. Maintaining a conversation today is of no interest to the mass consumer today. But except competent performance of tasks an important criterion is the convenience of using this tool. So, the next stage in the development of chatbots was active distribution of instant messengers. Statistically the number of users of WhatsApp in 2016 reached 1.6 billion users. "VK" is among the top three popular messengers in Russia. On the basis of this messenger interactive chatbot are created for completely different aims: information tasks, entertainment, sales, etc.

The Facebook company became the first company which introduced a chatbot in the messenger. Since then the development of the chatbots platforms has allowed companies to create their tools to attract customers and automate tasks.

Today it is customary to classify chatbots into the following categories:

- shopping;

- search by specified parameters;
- content consumption and entertainment;
- technical support.

The operation principle of the chatbot is presented in Figure 1.

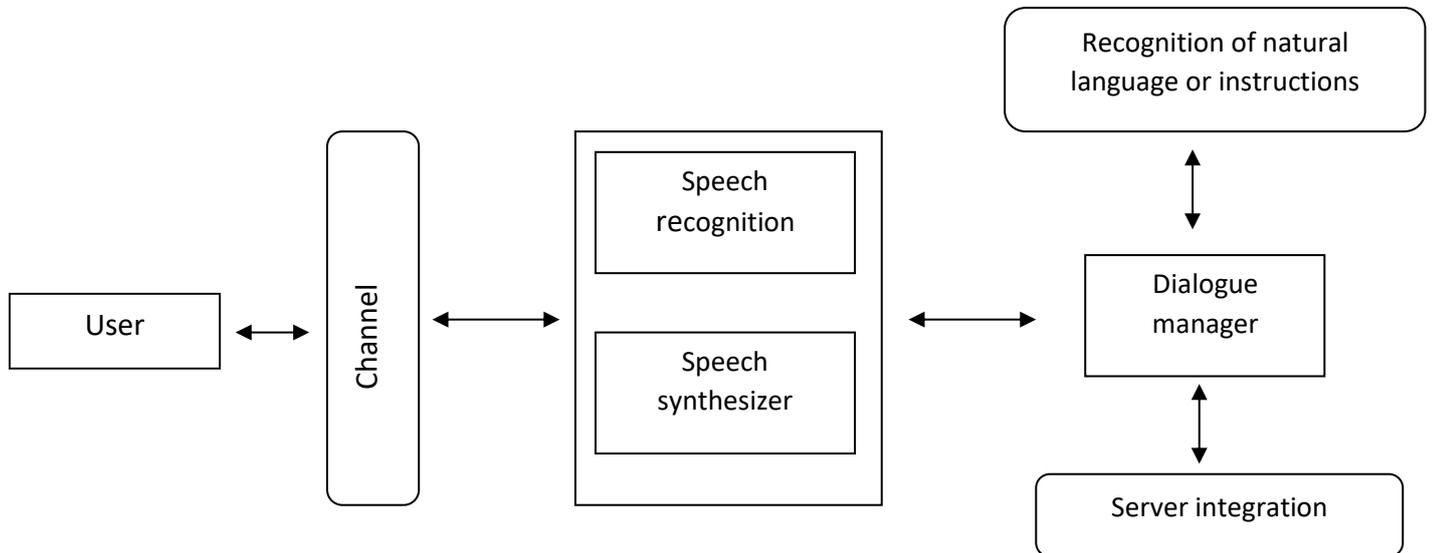


Figure 1. The operation principle of the chatbot

Processing the request is a search for the prepared instructions or analysis of the incoming text depending on the classification.

The response processing included both text generation and information gathering depending on the request.

In addition to responding to request, chatbots can contact users to provide or collect information.

Chatbot classification

There are two types of chatbots:

- chatbots that accept certain commands;
- chatbots that can analyze incoming information and extract meaning from the text.

The first type of a chatbots is used for narrowly targeted purposes and is not intended for the mass consumer. Certain commands pre-laid in algorithms to work. However, most chatbots work on this principle today.

The second category includes more complex programs. These chatbots use neural networks and deep training, so they are able to accept any users' requests, process them and issue answers. Thus, the chatbot is able to communicate in a natural language for person.

Neural networks imply mathematical models, as well as their software or hardware implementations, created according to the principle of organization and functioning of biological neural networks – networks of nerve cells of a living organism.

Natural Language Processing (NLP) – generalized direction of artificial intelligence, as well as mathematical linguistics. The problems of natural language synthesis and computer analysis are being studied.

Chatbots are classified by function:

1. Optimizer.

The most extensive category of chatbots. This type means the solution of tasks, at the same time optimizing information from the website or the application.

2. Highly specialized chatbots.

In this category the functionality is limited. For example, these are applications that only allow to edit a photo.

3. Proactive chatbots.

Task of this category is providing only necessary information and data, without disturbing the user often and unnecessary information.

4. Social chatbots.

They respond to requests based on the interaction of groups of people. This happens through the use of platforms for communication. This category of chatbots becomes more and more popular, despite simple functionality.

5. Interlocutor.

Often such chatbots are used in customer support. For the company, it is important that the client has a positive impression.

6. Personal assistant.

As a rule, it is the a big platform which has a set of functions.

Chatbots development perspective

The active distribution of instant messengers as a tool for easy and quick messaging began in 2015.

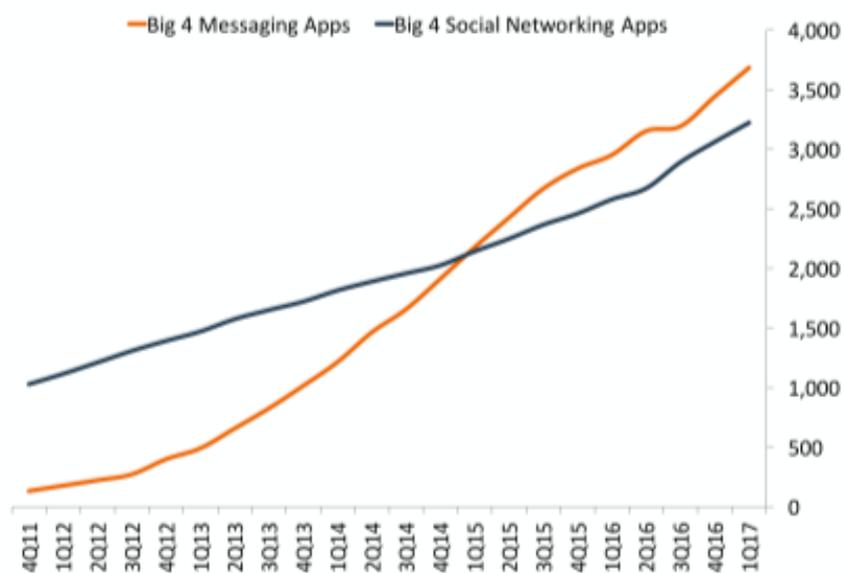


Figure 2. Monthly audience of messengers and social networks

Today chatbots on the basis of artificial intelligence have entered the life of a modern person very tightly. The development of chatbots in the future will be directed to technical improvement, as well as the problem will be solved the possibility to better understand the person’s language, that is, the transition to a more natural language. The range of applications of intelligent assistants will expand.

Talking about the high cost and complexity in the implementation and training of chatbots in natural language it is expected that in the near future they will only be used by major brands for which such developments are possible. Mass use will be focused on simple chatbots, which are based on patterns and commands. It is also expected chatbots will be able to most effectively solve typical problems of users, advise clients, give recommendations. They remember users data. When a client re-applies, he will not need to re-fill in personal data.

As for chatbots based on machine learning, their future in an ever closer approach to person thinking and communication, with the goal of accomplishing some more complicated tasks and processes than standard chat bots. The most

promising areas are the growth of speech recognition. This will allow users not to enter messages but to voice them, and the chatbot will understand their speech.

With the evolution of chatbots and the expansion of their technical and communication capabilities, the range of their applications will expand.

The leader in the implementation of chatbots is the banking sector, so in the future, it will set the main trends in the use of chat bots in communication with users and for internal analytics. Chatbots based on machine learning in the banking sector can develop and become financial advisers. They will analyze the structure of the income and expenses of the bank's clients, on the basis of which they will give it various tips on savings, investments, etc.

The introduction of chatbots is promising in the sphere of entertainment, because they represent a certain entertainment orientation already. Another sphere where the use of chatbots is promising and can be highly effective is public service.

So the following directions of chatbots evolution in the near future can be identified:

1. Mass distribution of narrow-featured chatbots based on the use of templates;
2. Expansion of the use of chatbots as consultants;
3. Using chatbots in consumer marketing research to collect and analyze data about them;
4. Using chatbots for development and sending personal offers to clients;
5. Using chatbots as advertising sites;
6. Expanding the capabilities and raising the intellectual level of chatbots based on machine learning, improving their understanding of speech and bringing their language closer to the natural;
7. Expansion sphere application of chatbots;
8. Use chatbots in Internet-trade.

The use of chatbots in corporate systems.

Importance of chatbots is an opportunity to unload perception, giving to the person the chance to interact habitually with subjects of communication in the interactive mode. The chatbot can ask a question and right there receive the answer which eliminates the need to independently search for this answer in arrays of information. The chatbot is similar to the information retrieval system, however, its value can be even higher. Since the process of communication with the chatbot is as interactive, as with a human being, this allows the user to be interested in continuing the communication and increase the degree of his involvement.

Today the large companies look for ways of introduction of intellectual assistants which are capable to carry out various routine office tasks. These are the examples of routine office tasks:

- staff recruitment;
- training of new employees;
- answers to questions;
- paperwork;
- planning of time and resources.

Robotization of business processes considerably accelerates processes due to automation of user interfaces. The robot is capable to perform the functions and tasks continuously. Complex corporate systems carry some problems. Such systems include storage of documentation, corporate mail, private offices with various functions. In addition, developers often do not provide for the adaptation of systems for mobile platforms, which creates inconvenience for employees located outside the office.

So, it is possible to define paramount functionality for chatbots:

- providing information;
- informing employees about news;
- mailing of reminders;
- sending electronic messages;
- filling of documents.

Using ready commands for so many functions is not a convenient solution. In this connection, it is possible to mark out ability of a chatbot to communicate in the language, natural to the person, as functionality. Safety is main criterion of introduction of the new tool in corporate networks for any large company. Safety means:

- lack of a possibility of leak of corporate data;
- lack of threat of breaking of internal network via this tool;
- reliability of the tool.

Chatbot experts with the level of knowledge of the system administrator have to be engaged in administration of the platform, to have initial knowledge in programming and system administration, to have initial knowledge in administration of databases.

The server hardware and all components of the platform for a chatbot have to run smoothly round the clock and daily, except for cases when there is a shutdown of power supply of the server hardware.

The platform for a chatbot should have function of soft stopping its work with closing of active connections, saving information, a stop of the relevant services and to transition to a waiting mode of switching off or reset. On completion of scheduled works, the platform by means of the provided function has to be returned to regular work, restoring work of the services and components.

Information security on the platform which processes data from a chatbot has to be realized both at the level of access to data, and at the level of information transfer on communication channels.

Corporate website or chatbot

Tasks which are solved by a corporate portal.

1. Organization of internal communications.

Portals are compared compare to social networks where employees can communicate in a habitual format. The most popular function – "a live tape" where

people obtain information in the form of posts. Also portals offer the chats which have to replace messengers.

2. Fast search information and documents.

Corporate portals solve a problem of separate storage of documents. Also they help with adaptation of new employees, providing fast access to the necessary information.

3. Automation of internal business processes.

Statement and tracking of performance of tasks, sending applications for compensatory leaves and holidays – all this can be carried out in a standard corporate portal. Besides, the task tracking is usually complemented with convenient calendars and a possibility of armor of negotiation there.

4. Integration with other internal systems.

Developers of portals promise ability to integrate with other systems of the company.

5. Organization of internal communications.

The chatbot is established in that messenger where employees communicate with each other and solve working problems. Thus the chatbot is naturally built in working communication environment, becoming one more member of collective (virtual). The information message sent from the face a chatbot becomes more personalized than simple news on the portal. According to stack overflow, 87% of messages in the messenger can be read within 3 minutes from the moment of sending.

6. Fast search information and documents.

The chatbot allows to look for information or documents in the set sources. For this purpose it is enough to execute in rather free form inquiry, with the indication of data which need to be found. The chatbot can instantly issue any information: from a birthday or contact of the colleague to the password to Wi-Fi. At the same time the chatbot can not only answer a question in a free form, but also allows to find information on a tree of categories independently.

7. Automation of internal business processes.

If the task is to automate complex business processes, then you can not do without the usual web interface. But there are tasks for which there is no need to use the corporate websites, for example, filling and sending the application for a holiday. For this purpose it is enough to execute inquiry at a chatbot, then to enter the required information and to wait for the notification about results of consideration.

8. Integration with other internal systems.

The chatbot can be integrated with any system not worse than an enterprise portal. All logic of work of a chatbot is implemented on the server therefore all opportunities of any software product are available to it. The messenger – is only the interface by means of which the chatbot communicates with the user. Therefore quite powerful tool is behind the seeming simplicity.

Problems of chatbots

Problems with chatbots arise in cases where, for any queries, the user must independently enter text into the dialogs. For a chatbot working on certain teams, the only problem is correct spelling of the command. For example, instead of “document”, the word “docment” is entered. In this case, the user must enter the request again.

However, for the chatbot understanding inquiries in a natural language, this mistake has to be only not a big obstacle in inquiry analysis. It leads to additional loading of the application. In this case it is necessary to apply mathematical algorithms to identification of similar words.

The main problem of natural language processing is language ambiguity. There are most different types of ambiguity: syntactic (structural), semantic ambiguity, etc. The applied systems of natural language processing have advantage before the general systems since work in narrow subject domains.

Nevertheless, creation of the systems having a possibility of communication in a natural language in wide areas is possible though so far results are far from satisfactory.

Приложение Б

Генеральному директору
ООО "Палтус"
Н.Л. Коровину
от экономиста
Г.С. Чурсина

ЗАЯВЛЕНИЕ

О предоставлении ежегодного оплачиваемого отпуска

Прошу предоставить мне очередной ежегодный оплачиваемый отпуск с 19.06.2019 на 14 дней.

15.04.2019

Г.С. Чурсин