**Федеральное государственное автономное образовательное учреждение высшего образования**

**«РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ»**

**(РУДН)**

Факультет физико-математических и естественных наук,

Кафедра прикладной информатики и теории вероятностей

02.03.02 «Фундаментальная информатика и Информационные Технологии»

**ОТЧЕТ**

**о прохождении учебной практики (научно-исследовательская работа (получение первичных навыков научно-исследовательской работы))**

(вид и наименование практики)

Ильин Никита Евгеньевич

(Ф.И.О. обучающегося)

Курс, группа 3, НФИбд-01-19

Место прохождения практики Отдел информационно-технологического обеспечения естественно-научных факультетов УИТО и СТС

полное наименование организации (предприятия)

Сроки прохождения с «04» 04.2022 г. по «15» 05.2022 г.

Руководители практики:

от РУДН доцент, к.н. Фомин М.Б.

Ф.И.О., должность

от организации доцент, к.н. Смирнов И.В.

Ф.И.О., должность

Оценка\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Москва 2022 г.

# Оглавление

[Оглавление 2](#__RefHeading___Toc476506714)

[Список сокращений 3](#__RefHeading___Toc476506715)

[Список основных обозначений 5](#__RefHeading___Toc476506716)

[Введение 6](#__RefHeading___Toc476506717)

[Основная часть 8](#__RefHeading___Toc476506718)

[Заключение 16](#__RefHeading___Toc476506719)

[Список использованных источников 18](#__RefHeading___Toc476506720)

# Список сокращений

**Русскоязычные** сокращения

Англоязычные сокращения

QA – Question Answer

Вопрос-ответ

# Список основных обозначений

# Введение

В рамках учебной практики была выполнена научно-исследовательская работа по теме «Вопросно-ответный поиск».

Согласно программе учебной практики направления подготовки 02.03.02 «Фундаментальная информатика и информационные технологии» целями практики являются:

* формирование профессиональных навыков в проведении научных исследований;
* формирование навыков использования современных научных методов для решения научных и практических задач;
* формирование практических навыков написания вспомогательных программных комплексов для проведения вычислительных экспериментов;
* формирование общекультурных, общепрофессиональный и профессиональных компетенций в соответствии с ОС ВО РУДН;
* формирование навыков оформления и представления результатов научного исследования;
* формирование навыков работы с источниками данных.

Там же определены задачи практики:

* формирование у студентов навыков в области изучения научной литературы и (или) научно-исследовательских проектов в соответствии с будущим профилем профессиональной деятельности и применения новых научных результатов;
* обучение правильному составлению научных обзоров и отчетов;
* формирование навыков решения конкретных научно-практических задач самостоятельно или в научном коллективе;
* обучение навыкам работы с прикладными комплексами программ для проведения вычислительных экспериментов;
* формирование способности разработки вспомогательных программных инструментов;
* обучение подготовке научных публикаций;
* формирование способности проводить научные исследования и получать новые научные и прикладные результаты.

Для достижения целей и решения поставленных задач в рамках учебной практики я выполнил обзор публикаций российских и международных научных изданий по теме практической работы, которая определена как «Вопросно-ответный поиск».

# Основная часть

В ходе данной работы я выполнил обзор публикаций российских и международных научных изданий по теме практической работы, которая определена как «Вопросно-ответный поиск».

1. Сопоставление синтаксических графов для улучшения модели вопросно-ответной системы [1]

В работе рассматривается один из методов улучшения модели вопросно-ответной системы – сопоставление синтаксических графов. Так как в настоящее время существует очень мало вопросно-ответных систем для русского языка, а разработанные на сегодняшний день системы часто выдают неудовлетворительные результаты, потому что дня него практически не используются семантические роли слов, а также системы способны отвечать только на простые работы, и в их работу заложены в основном алгоритмы статистического анализа текста, эффективность которых зачастую может оказаться недостаточной, становится очевидным факт, что необходимо повышать точность и быстродействие существующих систем .

Целью данной работы является разработка алгоритма, который позволяет автоматически находить ответы на вопросы к конкретным элементам текста. Исследование проводилось на текстах небольшого объема, состоящих из простых распространенных предложений на русском языке. Поиск ответа производился путем сравнения синтаксического графа вопроса, в котором вопросительное слово или словосочетание заменялось на маску, и синтаксических графов всех предложений текста. Наилучшие совпадения маски со словом или словосочетанием в тексте, отсортированные по релевантности, выдаются как ответ на вопрос. Разработанный алгоритм показал как высокое быстродействие (на предварительно размеченных текстах), так и высокую точность ответов.

Ввиду роста актуальности темы было необходимо разработать модель, с возможностью расширять вопросно-ответную систему, позволяющую дать точный ответ на поставленный вопрос, и имеющую высокое быстродействие. Расширяемость системы достигается за счет разделения системы на две независимых подсистемы, каждая из которых может быть улучшена, или заменена.

Моделью предложения является синтаксическое дерево, полученное в ходе автоматического анализа. С помощью морфологического анализа каждому слову присваивается грамматический признак, и с помощью определенного набора правил составляется дерево вероятности подчинения слов. Следующим шагом вероятности корректируются, путем учитывания контекста предложения, и замены местоимений антецедентами, процесс может повторяться несколько раз, пока не будет получен дерево с наибольшей вероятностью. Полученные синтаксические дерева позволяют проводить различные операции с предложениями, для выдачи максимально точного ответа.

В результате анализа модели система выдает, в виде конкретного слова или словосочетания, ответ на заданный вопрос. В ходе эксперимента, проведенного в работе, была показана высокая устойчивость алгоритмов, не менее 90%, в случае правильно разобранного вопроса и предложения, а в случае некачественного анализа – менее 35%. Это показывает, что решающим фактором верного ответа является качество автоматического анализа. Проведенные эксперименты помогли определить пути совершенствования системы

2. Диалоговый агент с обучаемым диалоговым менеджером [2]

В данной работе рассматривается создание диалогового агента в рамках соревнования по разработке диалоговых агентов. Разработанные ранее системы использовали подход, основанный на правилах для разработки стратегии ведения диалога. Такие системы трудны в поддержке, и требуют написания правила под каждую тематику диалога, также такие системы обладают низким качеством ответа. Соревнования по разработке диалоговых агентов общего назначения являются одним из путей привлечения людей к оценке качества, для создания более совершенных систем.

В данной работе была поставлена цель научить бота приветствовать пользователя, отвечать на вопросы о тексте, а также задавать их. Так как заданный текст может быть длинным, бот должен уметь автоматически реферировать его.

Часто пользователь имеет желание поговорить с ботом на общие темы, поэтому он также должен уметь это делать. Для поддержания таких диалогов были разработаны три навыка разговоров на общие темы. Разговор о тексте также может включать в себя вопросно-ответный сценарий, для чего были разработаны навыки генерации вопроса, ответа на вопрос, и проверки ответа на заданный вопрос. Бот был обучен на основе датасета SQuAD с помощью библиотеки OpenNMT. Для автоматического реферирования текста был также разработан соответствующий навык. Этот навык может помочь пользователю сэкономить время на вовлечение в беседу. Навык работает на основе датасета Gigaword на OpenNMT. При упоминании основной темы текста навык определения темы текста помогает с вовлечением пользователя в беседу. Навык определения был реализован с помощью библиотеки для тематического моделирования BigARTM. Дополнительно были реализованы более простые навыки, основанные на случайной отправке заготовленных фраз – навык приветствия и навык общих фраз.

Диалоговый менеджер состоит из двух классификаторов, которые принимают на вход контекст диалога или реплику пользователя, а на выход возвращают навык, выдающий ответ. Первый классификатор использует небольшую обучающую выборку, состоящую из нескольких фраз на навык, а второй – использует большую обучающую выборку, состоящую из нескольких фраз на навык и библиотеки fastText.

Анализ результатов показал, что наиболее популярным навыком стал разговорный, следовательно этот навык необходимо развивать, но несмотря на то, что навык реферирования вызывается всего один раз, это не значит, что он бесполезен, потому что этот навык вовлекает пользователя в беседу. Наличие нескольких разговорных навыков позволяет боту выдавать качественные реплики, и вовлекать пользователя в диалог.

3. Вопросно-ответный поиск в интеллектуальной поисковой системе Exactus [3]

Статья рассматривает отличительные особенности вопросно-ответного поиска. Показаны базовые принципы работы интеллектуальной поисковой системы Exactus. Вопросно-ответный поиск имеет существенные отличия от поиска, по ключевым словам, или его аналогов. Вопросы для поисковых систем чаще всего формируются в виде вопросительного предложения. Основной особенностью вопросно-ответного поиска заключается в поиске слова, словосочетания, или целого предложения, которое не содержится в явном виде в запросе. Вопросно-ответный поиск имеет другую цель в сравнении с обычным поиском. Его задача – найти фрагмент документа (веб страницы), содержащий точный ответ на заданный вопрос. Статья описывает ситуативно-реляционную модель текста, описывающую его семантику.

Если обратиться к теории коммуникативной грамматики русского языка, которая описывает подход к описанию русского синтаксиса, то можно выделить идею, заключающуюся в том, что синтаксис должен изучать именно осмысленную речь, а синтаксические правила должны учитывать категориальные значения слов, чтобы иметь возможность определять обобщенный смысл любой синтаксической конструкции.

Для построения модели предложения необходимо провести морфологический, синтаксический и семантический анализ текста. Для вопросно-ответного поиска анализ запроса, и анализ документов коллекции текстов. По тексту, для каждого предложения, строится семантическое дерево, которое является образом запроса и документов. Для реализации поиска необходимо вычислить релевантность результатов. Приоритетно именно сравнение синтаксем и связей, а не совпадение словоформ запроса.

В результате работы с помощью независимых экспертов удалось проверить работоспособность поисковых алгоритмов для дальнейшего развития системы.

4. Взвешенная погрешность — новая метрика для оценки качества валидации ответов в задаче вопросно-ответного поиска [4]

В данной работе рассмотрены существующие подходы к экспериментальной оценке качества вопросно-ответных систем, в частности модуля валидации ответов. Обоснован выбор новой метрики для выполнения экспериментов. Для оценки вопросно-ответной системы применяются некоторые метрики. Чтобы оценить валидацию ответа можно сравнить проходы системы в разных конфигурациях. Сравнив результаты этих прогонов, можно оценить вклад предлагаемой реализации модуля валидации ответов в качество вопросно-ответной системы. Процедура оценки вопросно-ответной системы очень трудоемка, так как требует работы нескольких асессоров, оценивающих результаты прогонов, но она позволяет выявить наиболее точные алгоритмы.

По результатам участия кампании РОМИП 2010, было принято решение исследовать подзадачу валидации ответа как задачу бинарной классификации. Была предложена новая метрика — взвешенная погрешность Eα, в отличие от традиционной F-меры учитывающая исходы true-negative, являющиеся важной категорией ответов для задачи валидации ответов.

5. Семантическая технология проектирования интеллектуальных вопросно-ответный систем [5]

Объектом рассмотрения в данной работе являются интеллектуальные вопросно-ответные системы, которые дают ответы на широкий спектр вопросов. Предложена модель данного класса систем, а также семантическая технология их проектирования. Поиск информации отнимает много времени, потому что существующие поисковые системы выдают большое количество ссылок, которые зачастую не имеют прямого отношения к теме.

Существующие подходы оптимизации информационного поиска в настоящее время лежат в плоскости разработки вопросно-ответных систем, в которых осуществляется сопоставление вопросов пользователей с требуемой информацией. Существующие системы AllQuest и AskNet Global Search ориентированы только на анализ и выявление семантических отношений между объектами предметной области в проиндексированных текстах. Данное обстоятельство накладывает следующие ограничения: нет возможности строго формально установить семантические отношения между объектами в тексте; невозможно сгенерировать ответ пользователю, когда такого ответа нет в проиндексированных текстах (т.е. в текущем информационном состоянии системы); не поддерживаются вопросы на выявление соответствий и аналогий между объектами и понятиями. Для устранения перечисленных ограничений требуется создание следующего поколения вопросно-ответных систем.

В работе были описаны основные понятия и принципы построения интеллектуальных вопросно-ответных систем. Прицнип многократного использования компонентов системы позволяет сократить сроки проектирования подобных систем за счет возможности использовать уже готовые типовые фрагменты баз знаний, операций и самих машин обработки знаний, пользовательских интерфейсов, оформленных в видекомпонентов интеллектуальной собственности.

Также в работе был описан язык запросов, который относится к семейству совместимых семантических языков. Язык запросов предназначен для формального описания поискового предписания вопросно-ответных систем, с целью удовлетворения информационной потребности пользователя. Объектами анализа языка вопросов являются классы вопросов в соответствии с семантической типологией вопросов.

Предложенная в работе семантическая технология предназначена для проектирования класса интеллектуальных вопросно-ответных систем, использующих в качестве формальной основы универсальный семантический код. Наличие совместимых семантических технологий проектирования баз знаний, машин обработки знаний и пользовательского интерфейса позволяют сократить сроки проектирования рассматриваемого класса систем за счет унификации способов представления знаний и легкой интегрируемости указанных компонентов.

6. Исследования и обзоры в вопросно-ответной системе (Research and reviews in question answering system) [6]

Вопросно-ответные системы – это автоматизированный подход к получению верных ответов на вопросы, заданные человеком на естественном языке. В данной работе рассматривается таксономия для характеристики вопросно-ответной системы, также кратко рассматриваются основные системы контроля качества ответов, и было проведено сравнение подходов, основанных на определенных особенностях системы контроля качества.

Главной проблемой для существующих систем контроля качества является правильное понимание вопросов на естественном языке, для выдачи верных ответов. Задача «ответ на вопрос» включает в себя методы искусственного интеллекта, обработки естественного языка, статистического анализа, сопоставления шаблонов, поиска и извлечения информации.

Были рассмотрены следующие подходы: лингвистический, статистический, составление шаблонов, на основе поверхностного шаблона, на основе шаблона. Все эти подходы хорошо работают в своих областях. Например, системы контроля качества, основанные на лингвистическом подходе, в основном, были построены на основе базы знаний для конкретной области. Статистический подход полезен для большого количества данных, система сможет давать ответы на сложные вопросы. Шаблонный подход вместо критического лингвистического анализа использует красноречие текстовых шаблонов. Даже несмотря на некоторые сбои, это удивительно эффективный метод, если использовать интернет, как источник данных. Однако, данному методу не хватает семантического понимания и рассуждений.

Так как все рассмотренные методы работают хорошо только в своих областях, стала очевидна необходимость создания универсального метода с гибридным подходом поиска. Были рассмотрены различные методы, такие как: MULDER, ASQA и IBM WATSON QA System.

В ходе работы получен вывод, что гибридные подходы обеспечивают лучшие результаты, но тем не менее очевидно, что в центре внимания останутся простые методы вопросно-ответных систем, потому что гибридные техники очень специфичны в работе.

7. Нейронная вопросно – ответная система для основных вопросов о подпрограммах (A Neural Question Answering System for Basic Questions about Subroutines) [7]

В работе рассматривается система вопросно – ответного поиска для базовых вопросов о подпрограммах. Цель работы сделать шаг к полноценной системе контроля качества для инженеров-программистов, необходимо продемонстрировать, как адаптировать эти технологии, чтобы можно было извлечь информацию из огромных наборов данных. Конечно, невозможно создать систему, которая будет отвечать на любые вопросы о программировании. Но идея состоит в том, чтобы создать систему, способную давать машинные ответы на информационные потребности программистов, которые настроены на контекст программного обеспечения и отдельные вопросы программистов. Для решения этой задачи необходимо решить несколько проблем, которые являются барьером в дальнейшем прогрессе. Успешная система должна не только ответить на узкую проблему, но и дать представление о том, как моделировать и извлекать функции из исходного кода, как интерпретировать информационные потребности программиста и как понимать словарь, который используют программисты, который отличается от общего использования слов.

Далее рассматривается справочная информация об интерактивных диалоговых системах и нейронном энкодере-декодере. Определяются методы выполнения работы. В ходе выполнения был создан набор данных для обучения, который позволил системе обучиться методам языка Java из более чем 10 тыс. проектов.

Затем была проведена оценка результатов, для которого были приглашены программисты. В ходе эксперимента выявлено, что модель может достоверно отвечать на поставленный вопрос, также стоит отметить, что система не предназначена для использования сама по себе, а предназначена в качестве компонента гораздо более широкой системы диалога.

8. SemBioNLQA: A semantic biomedical question answering system for retrieving exact and ideal answers to natural language questions [8]

Так как вопросно-ответные системы в биомедицинской области полноценно могут отвечать лишь только на ограниченное количество вопросов, а также требуют доработок для увеличения точности, и производительности. В работе была представлена семантическая биомедицинская система контроля качества под названием SemBioNLQA. Эта система способна обрабатывать запросы разных типов, на естественном языке.

Со временем усвоение информации в области биомедицины становится все сложнее из-за постоянно растущего объема, поэтому для поиска нужной информации может потребоваться много времени. Поэтому все чаще используются интеллектуальные методы анализа текста, однако в существующих системах размер ответа может быть велик, что также осложняет поиски точной информации, а некоторые вопросы вообще остаются без ответа. Поэтому важно иметь возможность получить отчет на любой тип вопроса на естественном языке. Вопросно-ответные системы, нацелены на предоставление точных ответов, путем автоматического анализа тысяч статей за несколько секунд. Поэтому целью данной работы является описать процесс проектирования и развития системы SemBioNLQA. Преимущества этой системы в том, что она решает проблемы предыдущих систем, такие как: Невозможность предоставить ответ для разных типов вопросов, невозможность предоставить идеальный ответ на естественном языке, проведение экспериментов с огромным количеством данных.

В работе были описаны методы, которые помогли достичь таких высоких показателей качества ответов. В качестве основных методов использовались: классификация вопросов и переформулирование запросов, извлечение информации из документов, поисковый проход, извлечение вопроса, вопрос да/нет, фактоидные вопросы, задание вопросов, обобщенные вопросы.

Результаты ручной оценки показали, что SemBioNLQA достигла лучших ответов, в сравнении с самой современной системой, а также смогла обеспечить конкурентоспособную альтернативу ей, помогая достичь точных ответов. Также, в будущем, рассматривается повышение производительности системы, в качестве основных направлений можно выделить: разные способы глубокого обучения, увеличение пар вопрос-ответ, улучшение работы с источниками.

9. Automatic question-answer pairs generation and question similarity mechanism in question answering system [9]

Вопросно-ответные системы играют важную роль в получении автоматических ответов на вопросы, с помощью информационных баз знаний. Данная работа включает в себя описание генерации, понимания и ответов на вопросы, преодолевая все ограничения систем. Первые вопросно-ответные системы появились еще в 1960-х годах, с того времени технологии поменялись, а количество информации возросло, поэтому доверие к системам повысилось. Системы научились выдавать ответ, максимально близкий к естественной речи. Но есть в этих системах и ограничения – это вопросы, на которые невозможно ответить. Если вопрос неверен, но связан с контекстом, то система может задать уточняющие вопросы. Точность и скорость ответов на вопросы от систем значительно превосходит человеческую.

Данная работа вносит определенный вклад в развитие вопросно-ответных систем, такие как:

* Автоматическая генерация возможных пар вопрос-ответ с учетом отрывка
* Ввод механизма подобия вопросов, на которые нет ответа, а также нерелевантные вопросы.
* Объединение системы генерации вопросов с системой ответов на вопросы, создав приложение «Автоматическая система генерации пар вопрос-ответ»

Автоматическая система генерации пар вопрос-ответ генерирует все возможные вопросы, и ответы на них. Такая система может применяться в разных областях. За счет введения механизма подобия вопросов система имитирует человеческие рассуждения, чтобы определить есть ли ответ на вопрос, существующие системы не умеют делать этого. Если вопрос не имеет ответа, то он не передается в вопросно-ответную систему. Также такая система не требует крупных вычислительных мощностей, поскольку в ней нет процесса обучения. Данный механизм можно включать в современные системы, за счет него модели могут сфокусироваться на ответах на вопрос, что повышает их производительность. А автоматически сгенерированные вопросно-ответные пары можно использовать, как набор данных для обучения вопросно-ответных систем.

10. Bangla Intelligence Question Answering System Based on Mathematics and Statistics

В работе рассматривается информационная система ответов на вопросы Bangla. Каждый хочет иметь систему, отвечающую на вопросы на своем родном языке, чтобы вопросы не ограничивались никакими правилами, или даже областью знаний. Новый подход соответствия потребностям пользователя заключается в проведении фактического анализа вопроса с лингвистической точки зрения и попытке понять, что на самом деле имеет в виду пользователь.

Система состоит из трех основных модулей: сбор данных, обработка информации и вопросов пользователей, а также установление взаимосвязи между ними. В процессе работы были внедрены математические и статические процедуры поиска информации, для предварительной обработки данных применили алгоритмы Хоббса, Edit Distance, и т.д., для упрощения предварительной обработки был разработан модуль, содержащий математические методы. Для разработки системы оценки качества были использовались пять основных компонентов, таких как лингвистическая предварительная обработка, распознавание сущностей, сопоставление элементов онтологии, обнаружение семантических связей, а также формулировка пар вопрос-ответ. Отличие от остальных систем состоит в том, что система работает на основе математики и статистики.

В процессе оценки исследований было показано, что разработанный чат-бот куда более современный, чем другие существующие. В будущем такая система может быть включена в образовательные, промышленные, деловые сферы, а также для личных задач и голосовых ассистентов. Дальнейшее развитие возможно с помощью алгоритмов глубокого обучения на рекуррентных нейронных сетях.

Исследования по построению и анализу математических моделей взаимодействия устройств в сетях Интернета вещей ведутся на кафедре ПИиТВ с 2014 года [1-10], и на сегодняшний день тема остается остро актуальной. Прогнозируемое увеличение объема трафика в беспроводных сетях создает высокие требования к пятому поколению мобильных систем (5G) [11,12]. Согласно ежегодному отчету компании Cisco «Наглядный индекс развития сетевых технологий» [11], ежегодный прирост трафика, создаваемый мобильными беспроводными устройствами, составляет 53% и по сравнению с 6,2 экзабайт в месяц (6,2∙1018) в 2016 г. достигнет 30,6 экзабайт в месяц к 2020 г. Одной из причин является рост числа пользовательских мобильных устройств – смартфонов, коммуникаторов, планшетов, носимых (wearable) устройств, таких как умные очки, браслеты, датчики слежения за физическим состоянием. Кроме того, прирост трафика связан с активным продвижением концепции интернета вещей [13], которая предусматривает автоматическое или с минимальным вовлечением человека производство и передачу данных межмашинного взаимодействия (machine-to-machine, M2M), собранных системами телеметрии и телеконтроля в жилищно-коммунальном хозяйстве (умный дом, умный город), сельском хозяйстве (сенсоры контроля технологических параметров в агрономии, датчики отслеживания и мониторинга в животноводстве), в интеллектуальных системах безопасности. По данным текущего отчета компании Ericsson «Ericsson Mobility Report» [12], в котором исследуются различные аспекты мобильной индустрии, на июнь 2016 г. на 5 млрд человек, проживающих в зоне действия беспроводных сетей, приходится 7,4 млрд мобильных подключений [12], а к 2020 г. эти цифры вырастут до 7,8 млрд и 11,6 млрд соответственно [11]. Таким образом, уже сегодня во многих странах число абонентов мобильной связи превышает численность населения, на душу населения приходится в среднем 1,5 мобильных устройства.

Беспроводные сети последующих поколений должны не только покрывать огромные территории с многомиллиардным населением, но и быть способны решить задачи социально-технологической эволюции следующего десятилетия, обеспечить разнообразие и масштабируемость услуг для пользователей. Сегодня действуют сети четвертого поколения 4G (International Mobile Telecommunications Advanced, IMT-Advanced), работающие на технологиях LTE Advanced (LTE-A) и WiMAX 2 (WMAN-Advanced), в России сотовые операторы осуществляют пользовательский доступ в Интернет на скорости передачи данных до 300 Мбит/с. Коммерческий запуск сетей стандарта 5G прогнозируется в 2020 г. В 2016 г. тестовые пилотные испытания мобильной сети пятого поколения на оборудовании компании Nokia показали пиковые скорости свыше 30 Гбит/с и более 1 млн одновременных подключений в соте [14], а в России в сети компании «Мегафон» на этом оборудовании удалось добиться скорости на уровне 5 Гбит/с [15]. Стандартизацией сетей пятого поколения 5G занимается созданная в 2015 г. оперативная группа по сетевым аспектам IMT-2020, которая опубликовала т. н. «дорожную карту» развития мобильной сети 5G, где обозначена задача разработки точных и подробных требований к новой сети для достижения сверхвысокой скорости передачи, ультранизкой задержки, высокой энергоэффективности и экологической безопасности.

Эти требования охватывают технологии радиоинтерфейсов, входящие в семейство IMT-2020, которые стали кандидатами для стандарта сетей пятого поколения, в том числе новые методы модуляции, кодирования, применение адаптивных антенных решеток с MIMO с узкой направленностью сигнала, а также ряд сетевых решений, обеспечивающих значительное увеличение производительности, включая использование малых (микро/пико/фемто) сот [16], ретрансляция через клиента (client-relays) [17], прямые соединения между конечными устройствами (D2D) [18]. Эти технологии основываются на повторном пространственном использовании частот, механизмы которого учитывают возникающую при этом повышенную интерференцию. Особенно заметно интерференция влияет на производительность беспроводной сети в гетерогенных сотовых сетях, где одновременное применение нескольких технологий беспроводного доступа сочетается с высокой плотностью пользователей. Универсальной метрикой определения производительности беспроводных систем является отношение сигнал/интерференция SIR в канале от ** к  радиосигнала [19], при этом интерференция зависит от *g*,  и от среды распространения сигнала. Отношение сигнал/интерференция влияет на показатели качества передачи в радиоканале, например, коэффициент ошибок на бит (BER), максимальную пропускную способность и спектральную эффективность канала связи, которые для услуг сетей последующих поколений определены международными стандартами.

Таким образом, задача нахождения отношения сигнал/интерференция для типичных сетевых конфигураций имеет особое значение, поскольку ее решение позволяет оценить применимость сценариев взаимодействия устройств в современных и перспективных беспроводных сетях. При анализе интерференции для моделирования расположения мобильных устройств традиционно применяются пространственные точечные процессы [20], которые позволяют для устройства-приемника в одной из точек процесса оценить искомую метрику – отношение сигнал/интерференция – с учетом интерференции от остальных устройств-передатчиков – соседних точек процесса. Так, в [1] был исследован пространственный пуассоновский точечный процесс (PPP), для которого известна формула расстояния до *n*-ближайшего соседа, а также процесс твердого ядра Маттерна (MHP), который позволяет учесть ограничения на расстояние внутри пары приемник – передатчик. Для этих процессов показана сложность получения распределения величины отношения сигнал/интерференция в замкнутой аналитической форме, поэтому актуальной является задача подбора аппроксимации для нахождения отношения сигнал/интерференция.

В статье [3] проблема нахождения отношения сигнал/интерференция исследуется для случая прямого взаимодействия устройств (D2D-взаимодействия) внутри помещений, например, в офисных зданиях или торгово-развлекательных центрах, где для обеспечения покрытия на каждом этаже размещается несколько точек беспроводного доступа (AP), имеющих относительно небольшую зону покрытия. При этом предполагается, что соседние точки располагаются в смежных помещениях. Особенностью этой задачи является фиксированное расположение передатчиков, что не позволяет при анализе интерференции воспользоваться в полной мере аппаратом стохастической геометрии – дисциплины, изучающей взаимоотношения между геометрией и теорией вероятностей, которая развилась из классической интегральной геометрии и задач о геометрической вероятности с привнесением идей и методов теории случайных процессов, в особенности теории точечных процессов. Для исследованного случая аналитические формулы, учитывающие интерференцию от нескольких передатчиков из смежных помещений, получены в предположении об идентичности основных характеристик интерференции для всех источников и являются развитием результатов для случая одного интерферирующего источника, известных из [2]. Также в [3] получен метод оценки плотности распределения с. в. отношения сигнал/интерференция с помощью аппроксимации мощности интерферирующего сигнала несколькими классическими распределениями, который учитывает потерю мощности при прохождении сигнала сквозь различные среды распространения (пространство внутри помещений, стены, межэтажные перекрытия). В статье [3] проведено сравнение аппроксимации для случая нормального и гиперэкспоненциального распределений с разработанной имитационной моделью, а также проведена оценка отношения сигнал/интерференция для различных материалов стен.

Также актуальность данной тематики можно показать на примере статьи «Обзор усовершенствований 3GPP для концентрации межмашинных взаимодействий» [21]. В данной статье представлена Технология eMTC, которая представляет ряд характерных свойств физического слоя, нацеленных на сокращение стоимости и потребления энергии пользовательского оборудования, а также на расширение сферы охвата, и в тоже время повторное использование большинства процедур физического слоя LTE. Технология eMTC UE может быть развёрнута в любом LTE, включающем узел B (eNB) для поддержки eMTC, а также может быть использован вместе с другими LTE UE с помощью того же узла eNB. Это позволяет развёртывать технологию eMTC с существующей инфраструктурой просто путём обновления программного обеспечения [21]. Статьи [21] и [22] дополняют друг друга. Таким образом актуальность данной тематики очевидна: При небольших изменениях UE и LTE можно добиться сокращения стоимости и потребления энергии, в то время, как зона охвата будет расширена.

Обзор основ концепции узкополосного Интернета вещей консорциума 3GPP [22]. В данной статье присутствует описание технической стороны NB-IoT. Основной причиной выбора данной статьи для дальнейшего использования ее в своей ВКР является

* описание схемы передачи данных и параметров развертывания NB-IoT
* описание физических каналов связи
* наглядная демонстрация распределения ресурсов.

Регулируемый множественный доступ в беспроводной сети умных вещей [23]. По мнению большинства специалистов, прошедший 2016 г. стал переломным при переходе к новой технологической платформе информатизации. Выбор аппаратно-программных средств на этапе внедрения новых технологий является не тривиальной задачей. Принимая соответствующие решения, следует учитывать

* диалектику взаимодействия технологий, реализующих информационные процессы;
* информационный объем автоматизируемых информационных процессов;
* возможность работы в реальном масштабе времени;
* ресурсные потребности сравниваемых платформ;
* инструментальные возможности обслуживания пользователей;
* проблемы информационной безопасности;
* надежность аппаратного и программного обеспечения.

Интернет вещей – это один из ключевых элементов новой платформы наряду с технологиями облачных вычислений больших данных, широкополосного мобильного доступа и наложенных сервисов [23].

В статье также описываются технологии передачи данных в интернете вещей такие, как

* персональные (Personal Area Network)
* локальные (Local Area Network)
* региональные (Metropolitan Area Network)
* глобальные (Wide Area Network)

Основным аспектом рассмотрения данной статьи, как вспомогательной литературы стало описание расчёта вероятностно – временных характеристик в беспроводной сенсорной сети, состоящей из двух подсетей (I и II): первого (высшего) и второго (низшего) приоритета соответственно. Предложенная процедура регулирования доступа УВ двух типов к эфирному каналу беспроводной сенсорной сети IoT, позволит учесть особенности обслуживания трафика, имеющего неоднородный характер. Также в статье представлена математическая модель беспроводной сенсорной сети с регулируемым синхронно-временным доступом к эфирной среде и найдены формулы для расчёта вероятностно-временных характеристик передачи в ней двух типов пакетов данных.

При расчёте вероятностно-временных характеристик применяются специальные программные пакеты, которые позволяют провести численный эксперимент и наглядно продемонстрировать его результаты. Для проведения численного эксперимента в своей ВКР я выберу язык программирования C++ и программу для работы с электронными таблицами Microsoft Excel. Возможно, также использую язык программирования R. Язык C++, по моему мнению - отличный язык программирования, которые поможет смоделировать реальную ситуацию. Если подключить необходимые библиотеки для этого языка программирования то возможно и графическое представление данных. При выборе компилятора C++ я остановлюсь на Dev C++- свободной интегрированной среде разработки приложений для языков программирования C/C++. В дистрибутив входит компилятор MinGW. Сам Dev-C++ написан на языке программирования Delphi и распространяется согласно открытому лицензионному соглашению GPL. Приложение Excel – программа для работы с электронными таблицами, созданная корпорацией Microsoft для Microsoft Windows, Windows NT и Mac OS, а также Android, iOS и Windows Phone. Она предоставляет возможности экономико-статистических расчетов, графические инструменты и, за исключением Excel 2008 под Mac OS X, язык макропрограммирования VBA. Приложение Microsoft Excel входит в состав Microsoft Office, и на сегодняшний день Excel является одним из наиболее популярных приложений в мире, потому что встроенные функции дают очень большие возможности без установки дополнительных пакетов. Именно поэтому он понадобится для расчёта вероятностных характеристик. Язык R – язык программирования для статистической обработки данных и работы с графикой, а также свободная программная среда вычислений с открытым исходным кодом в рамках проекта GNU. Программа R доступна под лицензией GNU GPL и распространяется в виде исходных кодов, а также откомпилированных приложений под ряд операционных систем: FreeBSD, Solaris и другие дистрибутивы Unix и Linux, Microsoft Windows, Mac OS X. В программе R используется интерфейс командной строки, хотя доступны и несколько графических интерфейсов пользователя, например, пакет RStudio. Язык R очень удобен как для просчета математических моделей, так и для работы с большим количеством данных, так как поддерживает широкий спектр статистических и численных методов и обладает хорошей расширяемостью с помощью пакетов. Пакеты представляют собой библиотеки для работы некоторых специфических функций или специальных областей применения. На сегодняшний момент доступно около 10210 пакетов.

# Заключение

За период практики, которую я проходил на кафедре прикладной информатики и теории вероятностей факультета физико-математических и естественных наук РУДН в научной лаборатории 5G Института прикладной математики и телекоммуникаций (ИПМиТ) РУДН, были достигнуты все цели и решены все поставленные задачи, определенные в программе учебной практики направления подготовки 02.04.02 «Фундаментальная информатика и информационные технологии» (см. введение отчета по практике).

При прохождении практики я разобрался с научной терминологией области исследований; научился осуществлять сбор, анализ и обработку данных, необходимых для решения профессиональных задач; собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям; осуществлять целенаправленный поиск информации о новейших научных и технологических достижениях в сети Интернет и из других источников; строить и анализировать математические модели объекта исследований; разрабатывать и отлаживать вспомогательные программные комплексы; проводить численный эксперимент; оформлять результаты своих исследований. Также я овладел необходимым математическим и программным аппаратом исследований; навыками математического моделирования, применения численных методов для выполнения необходимых расчетов и получения численных оценок по теме исследований.

В результате прохождения данной практики я приобрел следующие практические навыки, умения, универсальные и профессиональные компетенции:

* способностью к самоорганизации и к самообразованию;
* способностью решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности;
* способностью к самостоятельной научно-исследовательской работе;
* способностью находить, анализировать, реализовывать программно и использовать на практике математические алгоритмы, в том числе с применением современных вычислительных систем;
* способностью к определению общих форм и закономерностей отдельной предметной области;
* способностью математически корректно ставить естественнонаучные задачи, знание постановок классических задач математики;
* способностью строго доказать утверждение, сформулировать результат, увидеть следствия полученного результата;
* способностью публично представлять собственные и известные научные результаты;
* способностью использовать методы математического и алгоритмического моделирования при решении теоретических и прикладных задач;
* способностью передавать результат проведенных физико-математичес­ких и прикладных исследований в виде конкретных рекомендаций, выраженной в терминах предметной области изучавшегося явления;
* способностью представлять и адаптировать знания с учетом уровня аудитории;
* способностью к проведению методических и экспертных работ в области математики.

Выполненный во время проведения учебной практики обзор публикаций научных изданий по теме Вопросно-ответный поиск. позволит мне обосновать актуальность выбранной темы научно-исследовательской работы, а также более полно раскрыть постановку задачи и методы исследования при ее выполнении. Также в ходе учебной практики я подобрал программные средства для проведения численного эксперимента в научно-исследовательской работе.

# Список источников

[1] - А. В. Гашков, М. Н. Ельцова, Е. Л. Словикова – «Сопоставление синтаксических графов для улучшения модели вопросно-ответной системы» // Вестник ПНИПУ. Проблемы языкознания и педагогики № 1– 2021 г., с. 56 – 66

[2] – И. Ф. Юсупов, Ю. М. Куратов – «диалоговый агент с обучаемым диалоговым менеджером» // ТРУДЫ МФТИ. Том 12, № 4 – 2020 г., с. 106 – 120

[3] – И. А. Тихомиров – «вопросно-ответный поиск в интеллектуальной поисковой системе Exactus» // Труды РОМИП – 2006 г.

[4] – А. А. Соловьев – «взвешенная погрешность — новая метрика для оценки качества валидации ответов в задаче вопросно-ответного поиска» // Вестник МГТУ им. Н.Э. Баумана. Сер. “Приборостроение”. 2013. №1 – 2013, с. 58 – 64

[5] – С. А. Самодумкин – «Семантическая технология проектирования

интеллектуальных вопросно-ответных систем» // Доклады БГУИР – 2009 г., с. 67 - 72

[6] – Санджай К. Двиведия, Вайшали Сингхб – «Research and reviews in question answering system» // ScienceDirect: Procedia Technology 10 – 2013г., с. 417 – 424

[7] – Аакаш Бансал, Закари Эберхарт, Лингфэй Ву, Коллин Макмиллаy – «A Neural Question Answering System for Basic Questions about Subroutines» // 2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER), c. 70 – 71

[8] – Mourad Sarrouti, Said Ouatik El Alaouib – «SemBioNLQA: A semantic biomedical question answering system for retrieving exact and ideal answers to natural language questions» // Artificial Intelligence In Medicine 102 (2020) 101767, 2020 c. 1 – 16

[9] – Shivani G. Aithal, Abishek B. Rao, Sanjay Singh – Automatic question-answer pairs generation and question similarity mechanism in question answering system // Springer Nature 2021, c. 8485 – 8497

[10] – Md. Kowsher, M M Mahabubur Rahman, Sk Shohorab Ahmed, Nusrat Jahan Prottasha – Bangla Intelligence Question Answering System Based on Mathematics and Statistic // 22nd International Conference on Computer and Information Technology (ICCIT), 2019