|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство образования и науки Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА Системы обработки информации и управления

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

***К ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЕ***

***НА ТЕМУ:***

***«Модуль формирования метаграфа на основе текстового описания»***

Студент \_\_ИУ-5 84\_\_ **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_А.А.Фадеев\_\_\_**

(Группа) (Подпись, дата) (И.О.Фамилия)

Руководитель ВКР **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_В.Б.Тимофеев\_\_**

(Подпись, дата) (И.О.Фамилия)

Консультант **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_Ю.Е.Гапанюк\_\_**

(Подпись, дата) (И.О.Фамилия)

Нормоконтролер **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_Ю.Н.Кротов\_\_\_**

(Подпись, дата) (И.О.Фамилия)

2020 г.

РЕФЕРАТ

Расчётно-пояснительная записка квалификационной работы бакалавра содержит P страниц. С приложениями объем составляет N страниц. Работа включает в себя T таблиц и I иллюстраций. В процессе выполнения было использовано S источников.

Объектом разработки является программный модуль, формирующий метаграф на основе результатов синтаксического анализа русскоязычного текста.

После синтаксического анализа исходный текст преобразуется в метаграф. Он отражает синтаксическую структуру входной последовательности и хорошо подходит для дальнейшей обработки.

В процессе выполнения квалификационной работы бакалавра были изучены различные аспекты работы с естественным языком (ЕЯ) и её автоматизации.

Расчётно-пояснительная записка состоит из двух частей.

Постановка задач разработки рассматривает данную предметную область, процессы, подлежащие автоматизации при создании синтаксического анализатора и метаграфа, различные подходы и алгоритмы, которые для этого используются.

Конструкторско-технологическая часть включает выбор языка программирования для разработки программного изделия, проектирование, правил, базы данных, алгоритмов синтаксического анализа, формирования метаграфа и интерфейса пользователя.

В расчётно-пояснительную записку включены 3 приложения.

СОДЕРЖАНИЕ

[РЕФЕРАТ 2](#_Toc42096879)

[СОДЕРЖАНИЕ 3](#_Toc42096880)

[ВВЕДЕНИЕ 5](#_Toc42096881)

[1. ПОСТАНОВКА ЗАДАЧ РАЗРАБОТКИ 6](#_Toc42096882)

[1.1. Общетехническое обоснование разработки (Ваш вариант) 6](#_Toc42096883)

[1.1.1. Постановка цели работы 6](#_Toc42096884)

[1.1.2. Постановка задач 6](#_Toc42096885)

[1.1.3. Перечень процессов, подлежащих автоматизации 7](#_Toc42096886)

[1.2. Исходные данные 8](#_Toc42096887)

[1.2.1. Описание предметной области 8](#_Toc42096888)

[1.2.2. Выбор критериев качества 9](#_Toc42096889)

[1.2.3. Анализ аналогов и прототипов 10](#_Toc42096890)

[1.3. Анализ существующих алгоритмов и внедрение отличительных особенностей 10](#_Toc42096891)

[2. КОНСТРУКТОРСКО-ТЕХНОЛОГИЧЕСКАЯ ЧАСТЬ 11](#_Toc42096892)

[2.1. Выбор программных средств 11](#_Toc42096893)

[2.2. Выбор аппаратных средств 11](#_Toc42096894)

[2.3. Выбор языка разработки 11](#_Toc42096895)

[2.4. Структура программного изделия 11](#_Toc42096896)

[2.5. Проектирование структуры правил 11](#_Toc42096897)

[2.6. Проектирование модели базы данных 12](#_Toc42096898)

[2.7. Разработка программного продукта 27](#_Toc42096899)

[2.7.1. Разработка интерфейса пользователя 27](#_Toc42096900)

[2.7.2. Разработка алгоритмов синтаксического анализа текста 28](#_Toc42096901)

[2.7.3. Разработка структуры классов модуля построения метаграфа 28](#_Toc42096902)

[2.7.4. Разработка алгоритма преобразования результатов синтаксического анализа в метаграф 28](#_Toc42096903)

[2.8. Результаты опытной эксплуатации 28](#_Toc42096904)

[ЗАКЛЮЧЕНИЕ 29](#_Toc42096905)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 30](#_Toc42096906)

[ПРИЛОЖЕНИЕ А ГРАФИЧЕСКИЕ МАТЕРИАЛЫ 31](#_Toc42096907)

[ПРИЛОЖЕНИЕ В ТЕХНИЧЕСКОЕ ЗАДАНИЕ 34](#_Toc42096908)

[ПРИЛОЖЕНИЕ В ПРОГРАММА И МЕТОДИКА ИСПЫТАНИЙ 40](#_Toc42096909)

ВВЕДЕНИЕ

Квалификационная работа на тему «Формирование метаграфа на основе текстового описания» посвящена разработке модуля для построения метаграфа по результатам синтаксического анализа исходного русскоязычного текста.

В данной области множество научных работ и различных продуктов, дошедших до простого пользователя: электронные переводчики, распознаватели текста, поисковые системы, вопросно-ответные системы. Но почти все доступные сторонним пользователям системы не понимают слова так, как это делает человек, а лишь обрабатывают цепочки символов, используя статистику.

При работе с ЕЯ основными задачами являются разбор и анализ входного текста или синтез выходных предложений. Анализа текстов на ЕЯ состоит из 4-х этапов: морфологический, синтаксический, семантический и прагматический.

В данной квалификационной работе подробно рассматривается синтаксический этап разбора входного текста. Морфологический этап разбора проводится за счёт сторонних программных средств, о которых говорится в пункте 2.1 данной работы.

Морфологический анализ включает определение морфологических характеристик слова, таких как часть речи, род, число, падеж и т.д.

Синтаксический анализ, который содержит определение синтаксических ролей слов и словосочетаний в предложении. Результатом анализа является метаграф синтаксических зависимостей слов, который отражает структуру их взаимодействия в рамках одного предложения.

1. ПОСТАНОВКА ЗАДАЧ РАЗРАБОТКИ
   1. Общетехническое обоснование разработки (Ваш вариант)
      1. Постановка цели работы

В настоящее время существует множество систем, каким-либо образом анализирующих содержание текста. Однако почти все открытые для пользователей системы не анализируют текст как таковой, а лишь исследуют «важность» тех или иных слов внутри него, проводя максимально примитивный семантический анализ. Синтаксический анализ предваряет семантический, поэтому его роль в анализе текстов на ЕЯ не менее важна. Стоит разделять эти два этапа и по смыслу и алгоритмически.

В связи с этим цель работы формулируется следующим образом: предоставление возможности синтаксического анализа текста на русском языке и построения структуры, которая отражает синтаксическую структуру входной последовательности. А также предоставлении возможности консольной визуализации метаграфа-результата синтаксического разбора предложений текста.

* + 1. Постановка задач

В соответствии с целью данной работы, подразумевающей под собой создание синтаксического анализатора без каких-либо функций семантического анализатора, задачами проектирования становятся:

* Создание алгоритма фрагментации сложных предложений;
* Создание алгоритма поиска грамматической основы предложения;
* Создание структуры правил выделения синтаксических конструкций;
* Создание алгоритма применения правил для построения синтаксических групп (СГ) на фрагментах простых предложений в составе сложных;
* Формирование структуры синтаксического метаграфа;
* Создание алгоритма построения метаграфа;
* Разработка модуля консольной визуализации метаграфа.
  + 1. Перечень процессов, подлежащих автоматизации

При разработке модуля морфологического анализа текста следующие процессы подлежат автоматизации:

* Фрагментация текста на абзацы, абзацев - на предложения, а предложений - на лексемы (слова и знаки препинания);
* Формирование словаря, сохраняющего информацию о строении текста и морфологическую информацию о словах, то есть информацию о вариантах морфологического разбора каждой лексемы предложения.

При разработке модуля синтаксического анализа текста следующие процессы подлежат автоматизации:

* Формирование списка союзов каждого предложения;
* Фрагментация сложного предложения на простые с учётом информации о союзах и знаках препинания;
* Поиск грамматической основы каждой части предложения;
* Поиск вариантов цепочек словоформ следуя результатам морфологического разбора для лексем каждой части предложения;
* Применение правил к очередному варианту части предложения;
* Запоминание успешных попыток анализа части предложения;

При разработке модуля формирования метаграфа, как результата синтаксического разбора предложения:

* Формирование списка метаграфов на основе информации об абзацах, и положении предложений текста в абзацах, а абзацев в тексте.
  1. Исходные данные
     1. Описание предметной области

Предметной областью данной информационной системы является ххххххх хххххх хххххххх хххххххххххххх ххххххх хххххххх хххххххх ххх ххххх хххххх ххххххх ххххххххх ххххххххх ххххххххх ххххххх ххххххх ххххх хххх ххх ххххх хх ххххххххххх ххх ххххххх хххххх хххх хххххххххх ххххх ххххх хх хх хххххххххххх ххх ххххх хххххххххх хх ххххх хххххх ххх хххххххх ххххх ххх ххххх ххххххх.

В данную предметную область (рис.1) входят определенные сущности и взаимосвязи между ними. ххххххх хххххх хххххххх хххххххххххххх ххххххх хххххххх хххххххх ххх ххххх хххххх ххххххх ххххххххх ххххххххх ххххххххх ххххххх ххх ххххх хххххххххх хх ххххх хххххх ххх хххххххх ххххх ххх ххххх ххххххх. Графическое представление предметной области показано на рисунке 26.

Говоря о неживых объектах, то в системе могут быть представлены [1]: ххххххх хххххх хххххххх хххххххххххххх ххххххх хххххххх хххххххх ххх ххххх ххх ххххххх хххххх хххх хххххххххх ххххх ххххх хх хх хххххххххххх ххх ххххх хххххххххх хх ххххх хххххх ххх хххххххх ххххх ххх ххххх ххххххх. Количество сущностей увеличивается вместе с охватами все больших объемов информации [2] и со стремлением к достижению универсальности приложения.

* + 1. **Выбор критериев качества**

Для разрабатываемого программного изделия приоритетными являются следующие критерии качества [3]:

1. хх ххххххххххх ххх ххххххх хххххх [4];
2. хх ххххххххххх ххх ххххххх хххххх;
3. хх ххххххххххх ххх ххххххх хххххх;
4. хх ххххххххххх ххх ххххххх хххххх;
5. хх ххххххххххх ххх ххххххх хххххх;
6. ххх ххххххх хххххх.

Назначим весовые коэффициенты обозначенным критериям. Результаты приведены ниже в таблице 1.

**Таблица 1. Проранжированные критерии качества**

| **№**  **п/п** | **Название критерия** | **Весовой коэффициент** |
| --- | --- | --- |
| 1. | Возможность отправки тренировки | 4α |
| 2. | Отсутствие привязанности к определенному виду спорта | 3 α |
| 3. | Наличие банка знаний | 3 α |
| 4. | Среднее время загрузки (мс) | 2 α |
| 5. | Возможность доступа с разных устройств [5] | α |
| 6. | Отсутствие привязанности к определенному виду спорта | 3 α |
| 7. | Наличие банка знаний | 3 α |

* + 1. Анализ аналогов и прототипов

Общие требования к анализируемым системам:

Определим величину α, рассчитав значения весовых коэффициентов по формуле 1 [6].

, (1)

где *п* – общее количество весовых коэффициентов;

*ai* – весовые коэффициенты.

Подставив значения в формулу 1 получаем:

Проведём сравнительный анализ по методу взвешенных локальных критериев наиболее популярных мобильных приложений, которые соответствуют требованиям, приведенным ранее.

* 1. Анализ существующих алгоритмов и внедрение отличительных особенностей

Партия

1. КОНСТРУКТОРСКО-ТЕХНОЛОГИЧЕСКАЯ ЧАСТЬ
   1. Выбор программных средств

Партия

* 1. Выбор аппаратных средств

Партия

* 1. Выбор языка разработки

Партия

* 1. Структура программного изделия

Партия

* 1. Проектирование структуры правил

Партия

* 1. Проектирование модели базы данных

Для создания оптимальной базы данных необходимо пройти следующие три этапа:

* Концептуальное проектирование БД, т.е. построение инфологической модели БД;
* Логическое проектирование БД, т.е. построение даталогической модели БД;
* Физическое проектирование БД, т.е. построение самой БД средствами выбранной СУБД.

Основной задачей на этапе концептуального проектирования БД является выделение объектов данных, их свойств и связей между ними. Построение оптимальной инфологической модели позволит минимизировать возможность создания лишних, или отсутствия необходимых объектов и связей, замедляющих процесс обработки данных при работе разрабатываемого ПО. Основными понятиями на данном этапе являются: объекты, сущности, параметры, атрибуты.

Объект – элемент действительности, обладающий набором параметров, т.е. свойств. Отображение какого-либо объекта называется сущностью. Параметры объекта хранятся при сущности в виде её атрибутов.

В базе данных сущность становится таблицей. В таблице, как всегда, есть строки и столбцы. Одна строка называется записью. Запись - конкретный экземпляр данной сущности, обладающий тем же набором атрибутов, что и другие записи в таблице. Поля записи содержат значения атрибутов. Таким образом столбец - множество значений одного атрибута у всех записей из таблицы. Среди атрибутов выделяется один или несколько, которые называются первичным ключом, если для них выполняется условие уникальности, позволяющее однозначно определить каждую конкретную запись.

Связи отражают отношения между сущностями, и реализуются с помощью внешних ключей, т.е. полей первичных ключей, перешедших из других таблиц.

На этапе концептуального проектирования строится модель «Сущность-Связь», или ER-модель, или Инфологическая модель. Используются следующие графические средства: сущность, атрибут, связь.

По кардинальности связи классифицируют на три вида: «один-к-одному», «один-ко-многим», «многие-ко-многим». Связь «один к одному» строится тогда, когда для двух сущностей выполняется взаимное условие: один экземпляр одной сущности может быть связан с не более чем одним экземпляром другой сущности. Связь «один-ко-многим» строится, когда для двух сущностей выполняется условие: один экземпляр первой сущности может быть связан с несколькими экземплярами второй сущности, а один экземпляр второй сущности может быть связан с не более чем одним экземпляром первой сущности. Связь «многие-ко-многим» строится, когда для двух сущностей выполняется взаимное условие: один экземпляр одной сущности может быть связан с несколькими экземплярами другой сущности.

Каждой связи присваивается имя, которое отражает семантику взаимодействия экземпляров сущностей.

Прежде чем строить Инфологическую модель, необходимо построить таблицы сущностей будущей БД с указанием типов атрибутов и указанием ключевых атрибутов.

Ключевая сущность БД: «Правило». Она необходима для сохранения информации обо всех применяемых правилах, касающихся выделения синтаксических конструкций простого предложения. Именно эти правила сочетаемости словоформ применяются на этапе синтаксического анализа.

Сущность «Правило» представлена в таблице 1.

**Таблица 1. Сущность «Правило»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID правила | Целочисленный | Первичный (PK) |
| Название | Текстовый |  |
| Приоритет | Числовой |  |
| Флаг согласования рода | Логический |  |
| Флаг согласования числа | Логический |  |
| Флаг согласования падежа | Логический |  |

Ниже в таблицах 1-2 представлены две сущности, которые отвечают за сохранение информации о примененных правилах и тех словах, к которым эти правила были применены. В связи с выбранным алгоритмом синтаксического анализа, необходимо сохранить информацию о главной и подчиненной СГ.

Сущность «Главная синтаксическая группа» представлена в таблице 2.

**Таблица 2. Сущность «Главная синтаксическая группа»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID главной группы | Целочисленный | Первичный (PK) |
| Содержимое группы | Текстовый |  |
| Приоритет | Числовой |  |
| Номер предложения | Числовой |  |
| Номер части | Числовой |  |
| Номер варианта | Числовой |  |
| Номер в части | Числовой |  |

Сущность «Подчиненная синтаксическая группа» представлена в таблице 3.

**Таблица 3. Сущность «Подчиненная синтаксическая группа»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID подчиненной группы | Целочисленный | Первичный (PK) |
| Содержимое группы | Текстовый |  |
| Приоритет | Числовой |  |
| Номер предложения | Числовой |  |
| Номер части | Числовой |  |
| Номер варианта | Числовой |  |
| Номер в части | Числовой |  |

Сущность «Тип входной синтаксической группы» представлена в таблице 4. Она описывает два возможных типа: «подчиненный» и «главный», в соответствии с которыми идёт управление двумя СГ, которые связываются по какому-либо правилу. Таким образом типы 1-й и 2-й СГ определяется правилом.

**Таблица 4. Сущность «Тип входной синтаксической группы»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID типа входной синтаксической группы | Целочисленный | Первичный (PK) |
| Название типа входной синтаксической группы | Текстовый |  |

Сущность «Вид синтаксической группы» представлена в таблице 5. Данная сущность унифицирует синтаксические группы по видовому признаку, а именно их главные слова, благодаря чему делается возможным применение правила к двум СГ.

**Таблица 5. Сущность «Вид синтаксической группы»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID вида синтаксической группы | Целочисленный | Первичный (PK) |
| Название вида синтаксической группы | Текстовый |  |

Сущность «Часть речи» представлена в таблице 6. Каждое слово - представитель определенной части речи русского языка, поэтому эта характеристика - обязательная связь между русским языком и принципом СГ.

**Таблица 6. Сущность «Часть речи»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID части речи | Целочисленный | Первичный (PK) |
| Название части речи | Текстовый |  |

Сущность «Граммема» представлена в таблице 7.

**Таблица 7. Сущность «Граммема»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID граммемы | Целочисленный | Первичный (PK) |
| Содержание граммемы | Текстовый |  |

Сущность «Класс синтаксической группы» представлена в таблице 8.

**Таблица 8. Сущность «**Класс синтаксической группы**»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID класса | Целочисленный | Первичный (PK) |
| Название класса | Текстовый |  |

Сущность «Член предложения» представлена в таблице 9. Данная таблица дает информацию о том, какой член предложения реализуется той или иной СГ.

**Таблица 9. Сущность «Член предложения»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID члена предложения | Целочисленный | Первичный (PK) |
| Название члена предложения | Текстовый |  |

Ниже представлены три сущности, соответствующие трём видам граммем русского языка: род, число и падеж, в соответствии с которыми определяется лексически правильное взаимодействие слов.

Сущность «Род» представлена в таблице 10.

**Таблица 10. Сущность «Род»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID рода | Целочисленный | Первичный (PK) |
| Название рода | Текстовый |  |

Сущность «Число» представлена в таблице 11.

**Таблица 11. Сущность «Число»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID числа | Целочисленный | Первичный (PK) |
| Название числа | Текстовый |  |

Сущность «Падеж» представлена в таблице 12.

**Таблица 12. Сущность «Падеж»**

| **Логическое имя атрибута** | **Тип данных** | **Ключ** |
| --- | --- | --- |
| ID падежа | Целочисленный | Первичный (PK) |
| Название падежа | Текстовый |  |

Ключевые атрибуты, показанные в таблицах 1-12, на Инфологической модели БД выделяются подчёркиванием, но прежде, чем строить эту модель, необходимо определить связи между сущностями БД. А на основе ключевых атрибутов в дальнейшем будут строится первичные ключи в таблицах Даталогической модели.

Преобразование ER-модели в реляционную модель БД происходит с учётом типа связи:

* Связь «1:1». Для реализации данного типа связи требуется от одной до двух таблиц в зависимости от того, обязательно ли поле.
* Связь «1:М». Связь может быть реализована двумя или тремя таблицами.
* Связь «М:1». Также необходимо две или три таблицы.
* Связь «М:N». Такая связь требует обязательного построения третей таблицы – одну для каждой сущности и одну для связи.

Первичный ключ должен удовлетворять ряду требований:

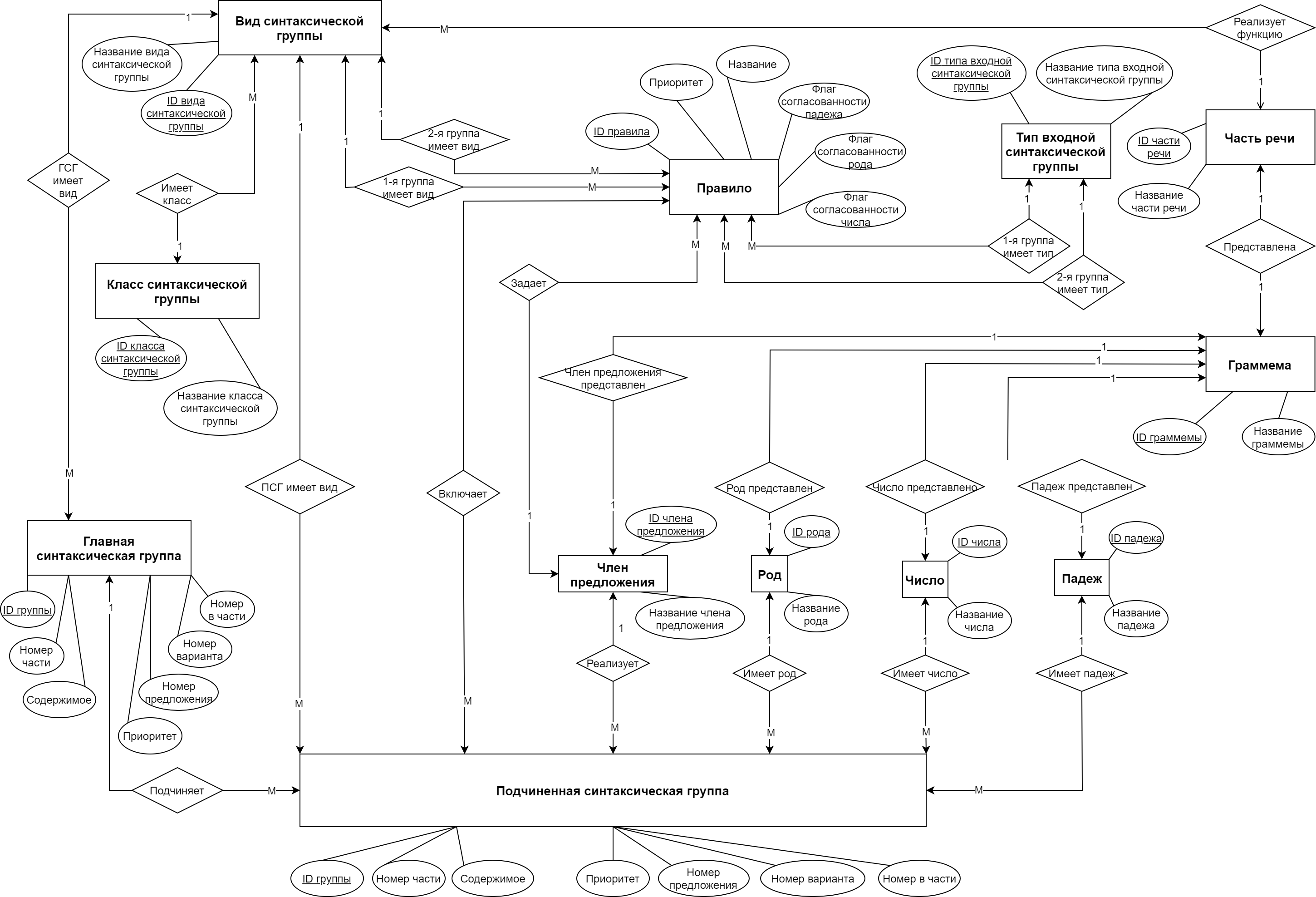
* Уникальность. Два экземпляра не должны иметь одинаковых значений ключа.
* Минимум атрибутов. Составной первичный ключ (группа атрибутов) не должен содержать ни одного атрибута, удаление которого не приводило бы к утрате уникальности.

Внешний ключ (Foreign Key) появляется в процессе установления связей между сущностями. Это описательный атрибут, который назначается в дополнение к идентифицирующему атрибуту и служит не для идентификации этого уникального экземпляра, а для выделения набора интересующих тех объектов по каким-либо внешним признакам, привязанным к этим объектам из других таблиц.

С учётом описания сущностей формируются следующие связи:

* «Род представлен»: связь между сущностями «Граммема» и «Род» типа «один-к-одному»
* «Число представлено»: связь между сущностями «Граммема» и «Число» типа «один-к-одному»
* «Падеж представлен»: связь между сущностями «Граммема» и «Падеж» типа «один-к-одному»
* «Часть речи представлена»: связь между сущностями «Граммема» и «Часть речи» типа «один-к-одному»
* «Член предложения представлен»: связь между сущностями «Граммема» и «Член предложения» типа «один-к-одному»
* «Реализует функцию»: связь между сущностями «Часть речи» и «Вид синтаксической группы» типа «один-ко-многим»
* «1-я группа имеет тип»: связь между сущностями «Тип входной синтаксической группы» и «Правило» типа «один-ко-многим»
* «2-я группа имеет тип»: связь между сущностями «Тип входной синтаксической группы» и «Правило» типа «один-ко-многим»
* «1-я группа имеет вид»: связь между сущностями «Вид синтаксической группы» и «Правило» типа «один-ко-многим»
* «2-я группа имеет вид»: связь между сущностями «Вид синтаксической группы» и «Правило» типа «один-ко-многим»
* «Имеет класс»: связь между сущностями «Класс синтаксической группы» и «Вид синтаксической группы» типа «один-ко-многим»
* «ГСГ имеет вид»: связь между сущностями «Вид синтаксической группы» и «Главная синтаксическая группа» типа «один-ко-многим» (ГСГ - Главная Синтаксическая Группа)
* «ПСГ имеет вид»: связь между сущностями «Вид синтаксической группы» и «Подчиненная синтаксическая группа» типа «один-ко-многим» (ПСГ - Подчиненная Синтаксическая Группа)
* «Определена»: связь между сущностями «Правило» и «Подчиненная синтаксическая группа» типа «один-ко-многим»
* «Подчиняется»: связь между сущностями «Вид синтаксической группы» и «Подчиненная синтаксическая группа» типа «один-ко-многим»
* «Реализует»: связь между сущностями «Член предложения» и «Подчиненная синтаксическая группа» типа «один-ко-многим»
* «Задает»: связь между сущностями «Член предложения» и «Правило» типа «один-ко-многим»
* «Имеет род»: связь между сущностями «Род» и «Подчиненная синтаксическая группа» типа «один-ко-многим»
* «Имеет число»: связь между сущностями «Число» и «Подчиненная синтаксическая группа» типа «один-ко-многим»
* «Имеет падеж»: связь между сущностями «Падеж» и «Подчиненная синтаксическая группа» типа «один-ко-многим».

На рисунке 1 представлена Инфологическая модель БД.

Рисунок 1. Схема Инфологической модели.

Под даталогической понимается модель, отражающая логические взаимосвязи между элементами данных безотносительно их содержания и физической организации. Даталогическое проектирование включает построение БД в виде таблиц. Особенности выбранной СУБД также влияют на проеткирование даталогической модели, например, при описании названий типов данных, их размера. Результатом данной стадии проектирования БД является построение датологической модели в виде набора таблиц, отражающих основные свойства и взаимодействия сущностей предметной области.

Таблица содержит название и набор полей, которые характеризуются типом данных. Поле может содержать информативные данных стандартных типов данных (целый, с плавающей точкой, строковый и т.д.) или быть ссылкой на другую таблицу – внешним ключом.

При добавлении информации в БД в таблицы вставляются записи, для которых каждое поле принимает значение заданного типа.

Проектированию базы данных уделяется особое внимание по причине того, что внесение изменений в структуру при наличии данных влечет дополнительные сложности, объясняющиеся необходимостью специальными методами перезаписывать и изменять все имеющиеся данные.

Данная предметная область облегчает построение Базы данных, поскольку ее данные используют лишь алгоритмы синтаксического анализатора. Она не будет содержать конфиденциальных или сколь угодно уникальных данных, поэтому БД проекта не нуждается в дополнительных средствах защиты данных, не предусмотренных базовыми настройками выбранной СУБД.

Модель «Rule» представлена в таблице 13.

**Таблица 13. Модель «Rule»**

| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID правила | ID | INT(11) | Auto Incremental | Primary Key |
| Название правила | Name | VARCHAR(45) | NULL |  |
| Приоритет | Priority | INT(11) | NULL |  |
| Флаг согласования рода | Gender consistency | TINYINT(4) | NULL |  |
| Флаг согласования числа | Number consistency | TINYINT(4) | NULL |  |
| Флаг согласования падежа | Case consistency | TINYINT(4) | NULL |  |
| Вид первой СГ | 1G Syntax kind ID | INT(11) | NULL | Foreign Key |
| Вид второй СГ | 1G Syntax kind ID | INT(11) | NULL | Foreign Key |
| Тип первой СГ | 1G Syntax kind ID | INT(11) | NULL | Foreign Key |
| Тип второй СГ | 1G Syntax kind ID | INT(11) | NULL | Foreign Key |
| Член предложения подчиненной группы | 1G SG attribute ID | INT(11) | NULL | Foreign Key |

Модель «Syntax kind» представлена в таблице 14.

**Таблица 14. Модель «Syntax kind»**

| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID вида СГ | ID | INT(11) | Auto Incremental | Primary Key |
| Название вида СГ | Name | VARCHAR(45) | NULL |  |
| Части речи | Part of speech ID | INT(11) | NULL | Foreign Key |
| Класс СГ | Syntax class ID | INT(11) | NULL | Foreign Key |

Модель «Input syntax type» представлена в таблице 15.

**Таблица 15. Модель «Input syntax type»**

| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID правила | ID | INT(11) | Auto Incremental | Primary Key |
| Название правила | Name | VARCHAR(45) | NULL |  |

Модель «Attribute» представлена в таблице 16.

**Таблица 16. Модель «Attribute»**

| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID члена предложения | ID | INT(11) | Auto Incremental | Primary Key |
| Название члена предложения | Name | VARCHAR(45) | NULL |  |

Модель «Syntax class» представлена в таблице 17.

**Таблица 17. Модель «Syntax class»**

| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID класса СГ | ID | INT(11) | Auto Incremental | Primary Key |
| Название класса СГ | Name | VARCHAR(45) | NULL |  |

Модель «Grammeme» представлена в таблице 18.

**Таблица 18. Модель «Grammeme»**

| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID граммемы | ID | INT(11) | Auto Incremental | Primary Key |
| Название граммемы | Name | VARCHAR(45) | NULL |  |

Модель «Gender» представлена в таблице 19.

**Таблица 19. Модель «Gender»**

| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID рода | ID | INT(11) | Auto Incremental | Primary Key |
| Название рода | Name | VARCHAR(45) | NULL |  |
| Граммема | Grammeme ID | INT(11) | NULL | Foreign Key |

Модель «Case» представлена в таблице 20.

**Таблица 20. Модель «Case»**

| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID падежа | ID | INT(11) | Auto Incremental | Primary Key |
| Название падежа | Name | VARCHAR(45) | NULL |  |
| Граммема | Grammeme ID | INT(11) | NULL | Foreign Key |

Модель «Number» представлена в таблице 21.

**Таблица 21. Модель «Number»**

| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID числа | ID | INT(11) | Auto Incremental | Primary Key |
| Название числа | Name | VARCHAR(45) | NULL |  |
| Граммема | Grammeme ID | INT(11) | NULL | Foreign Key |

Модель «Part of speech» представлена в таблице 22.

**Таблица 22. Модель «Part of speech»**

| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID часть речи | ID | INT(11) | Auto Incremental | Primary Key |
| Название части речи | Name | VARCHAR(45) | NULL |  |
| Граммема | Grammeme ID | INT(11) | NULL | Foreign Key |

Модель «Main syntax group» представлена в таблице 23.

**Таблица 23. Модель «Main syntax group»**

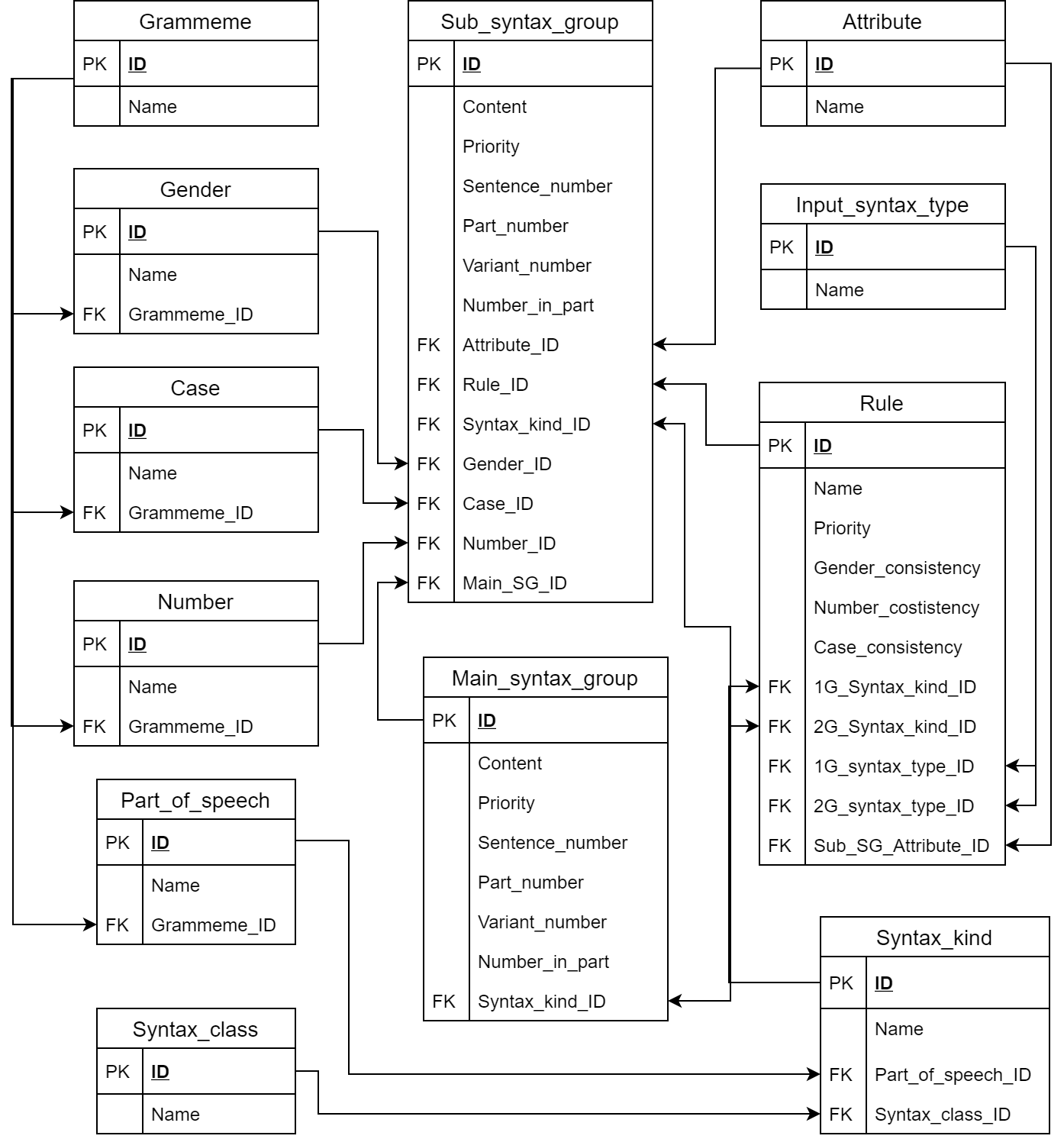
| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID группы | ID | INT(11) | Auto Incremental | Primary Key |
| Содержимое | Content | VARCHAR(45) | NULL |  |
| Приоритет | Priority | INT(11) | NULL |  |
| Номер предложения | Sentence number | INT(11) | NULL |  |
| Номер части | Part number | INT(11) | NULL |  |
| Номер варианта | Variant number | INT(11) | NULL |  |
| Номер в части | Number in part | INT(11) | NULL |  |
| Вид СГ | Syntax kind ID | INT(11) | NULL | Foreign Key |

Модель «Sub syntax group» представлена в таблице 24.

**Таблица 24. Модель «Sub syntax group»**

| **Логическое имя атрибута** | **Поле** | **Тип данных** | **Default** | **Ключ** |
| --- | --- | --- | --- | --- |
| ID группы | ID | INT(11) | Auto Incremental | Primary Key |
| Содержимое | Content | VARCHAR(45) | NULL |  |
| Приоритет | Priority | INT(11) | NULL |  |
| Номер предложения | Sentence number | INT(11) | NULL |  |
| Номер части | Part number | INT(11) | NULL |  |
| Номер варианта | Variant number | INT(11) | NULL |  |
| Номер в части | Number in part | INT(11) | NULL |  |
| Член предложения | Attribute ID | INT(11) | NULL | Foreign Key |
| Правило | Rule ID | INT(11) | NULL | Foreign Key |
| Вид СГ | Syntax kind ID | INT(11) | NULL | Foreign Key |
| Род | Gender ID | INT(11) | NULL | Foreign Key |
| Число | Number ID | INT(11) | NULL | Foreign Key |
| Падеж | Case ID | INT(11) | NULL | Foreign Key |
| Главная СГ | Main SG ID | INT(11) | NULL | Foreign Key |

Связи даталогической модели формируются на основе предыдущих этапов проектирования. На рисунке 2 представлена итоговая схема Даталогической модели.

Рисунок 2. Схема даталогической модели

* 1. Разработка программного продукта
     1. Разработка интерфейса пользователя

Партия

* + 1. Разработка алгоритмов синтаксического анализа текста

Партия

* + 1. Разработка структуры классов модуля построения метаграфа

Партия

* + 1. Разработка алгоритма преобразования результатов синтаксического анализа в метаграф

Партия

* 1. Результаты опытной эксплуатации

Общие требования к анализируемым системам:

ЗАКЛЮЧЕНИЕ

При разработке информационной системы планирования и учета личных тренировок были получены следующие результаты:

* Была изучена предметная область, связанная ххх ххххх хххххх ххххххх хххххххх хххххххх ххх ххххх хххххх ххххххх ххххххххх ххххххххх ххххххххх ххххххх ххххххх ххххх хххх ххх ххххх хх ххххххххххх ххх ххххххх хххххх хххх хххххххххх ххххх ххххх хх хх хххххххххххх ххх ххххх хххххххххх хх ххххх хххххх ххх хххххххх ххххх ххх ххххх ххххххх.
* Были сформулированы требования ххх ххххх хххххх ххххххх хххххххх хххххххх ххх ххххх хххххх ххххххх ххххххххх ххххххххх ххххххххх ххххххх ххххххх.
* Спроектирована ххх ххххх хххххх ххххххх хххххххх хххххххх ххх ххххх хххххх ххххххх ххххххххх ххххххххх ххххххххх ххххххх ххххххх ххххх.
* Разработана ххх ххххх хххххх ххххххх хххххххх хххххххх ххх ххххх хххххх ххххххх ххххххххх ххххххххх ххххххххх ххххххх ххххххх ххххх хххх ххх.

Полученное информационно-программное изделие имеет возможности расширения за счет подключения новых платформ к серверному приложению. Это позволит системе привлечь новых пользователей и оставаться конкурентно способной среди аналогов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Волкова, И. А. Введение в компьютерную лингвистику. Практические аспекты создания лингвистических процессоров: Учебное пособие для студентов ВМиК МГУ / И. А. Волкова - Москва: Издательский отдел факультета ВМиК МГУ, 2006. - 43 с.
2. Прикладная и компьютерная лингвистика / ред. Николаев И. С., Митренина О. В., Ландо Т. М. - 2-е изд., - Москва: Издательская группа URSS, 2017. - 320 с.
3. Гапанюк, Ю. Е. Конспект лекций по спецкурсу «Гибридные интеллектуальные информационные системы на основе метаграфового подхода» / Ю. Е. Гапанюк - Москва: Издательство «Спутник +», 2018. - 56 с.
4. Тузов, В. А. Компьютерная семантика русского языка / В. А. Тузов - СПб.: Изд-во СПбГУ, 2003. - 391 с.
5. Мартынов, В. В. Основы семантического кодирования. Опыт Представления и преобразования знаний / В. В. Мартынов. - Минск: ЕГУ, 2001 г. - 140 с.
6. [Электронный ресурс] - 2009-2020 г. - Режим доступа: http://opencorpora.org/dict.php?act=gram, свободный.
7. [Электронный ресурс] / М. Коробов - 2015-2020 г. - Режим доступа: https://pymorphy2.readthedocs.io/en/latest/index.html, свободный.

ПРИЛОЖЕНИЕ А   
ГРАФИЧЕСКИЕ МАТЕРИАЛЫ

1. Цели и задачи

2. Предметная область

3. Сравнение с аналогами

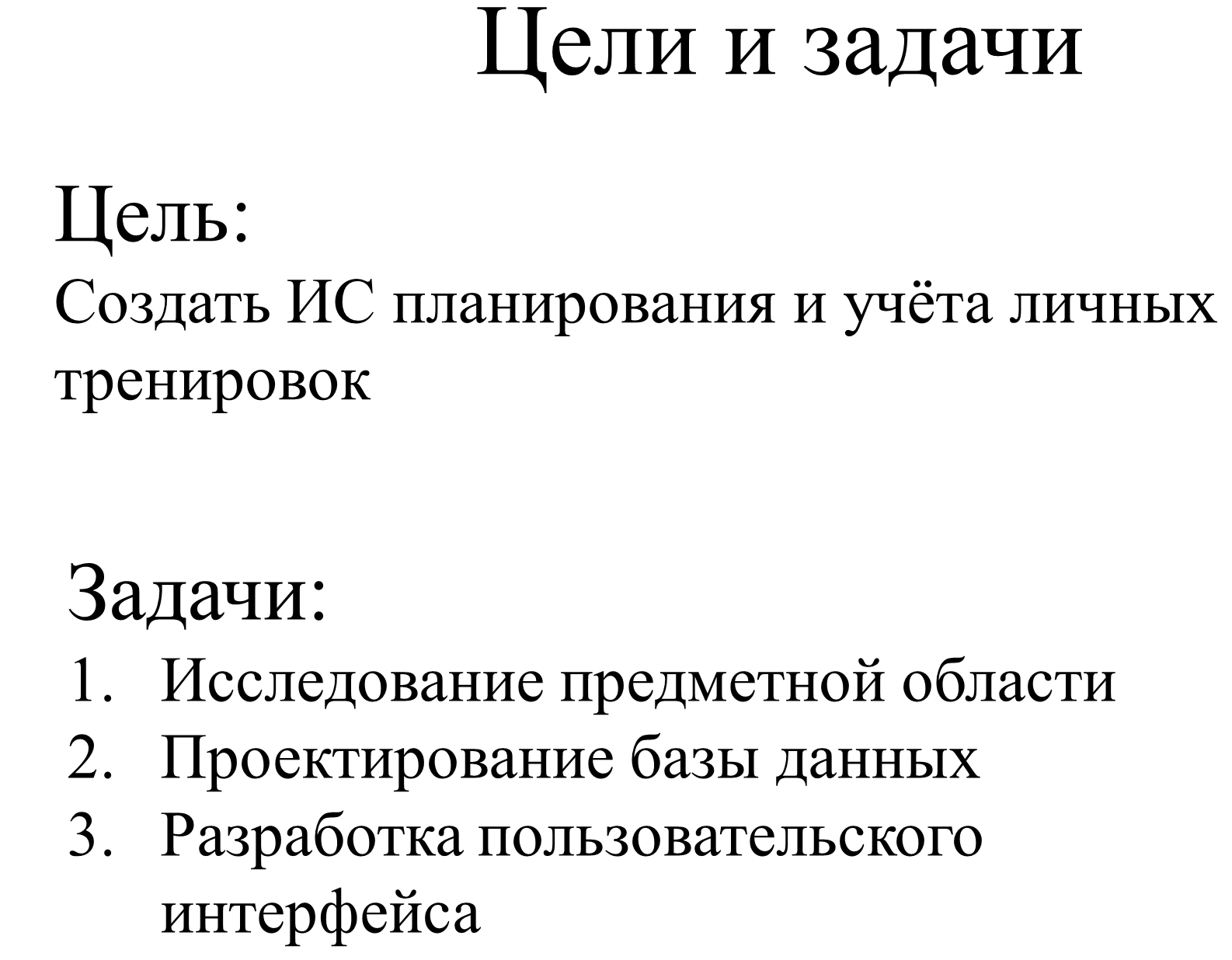
4. Архитектура системы

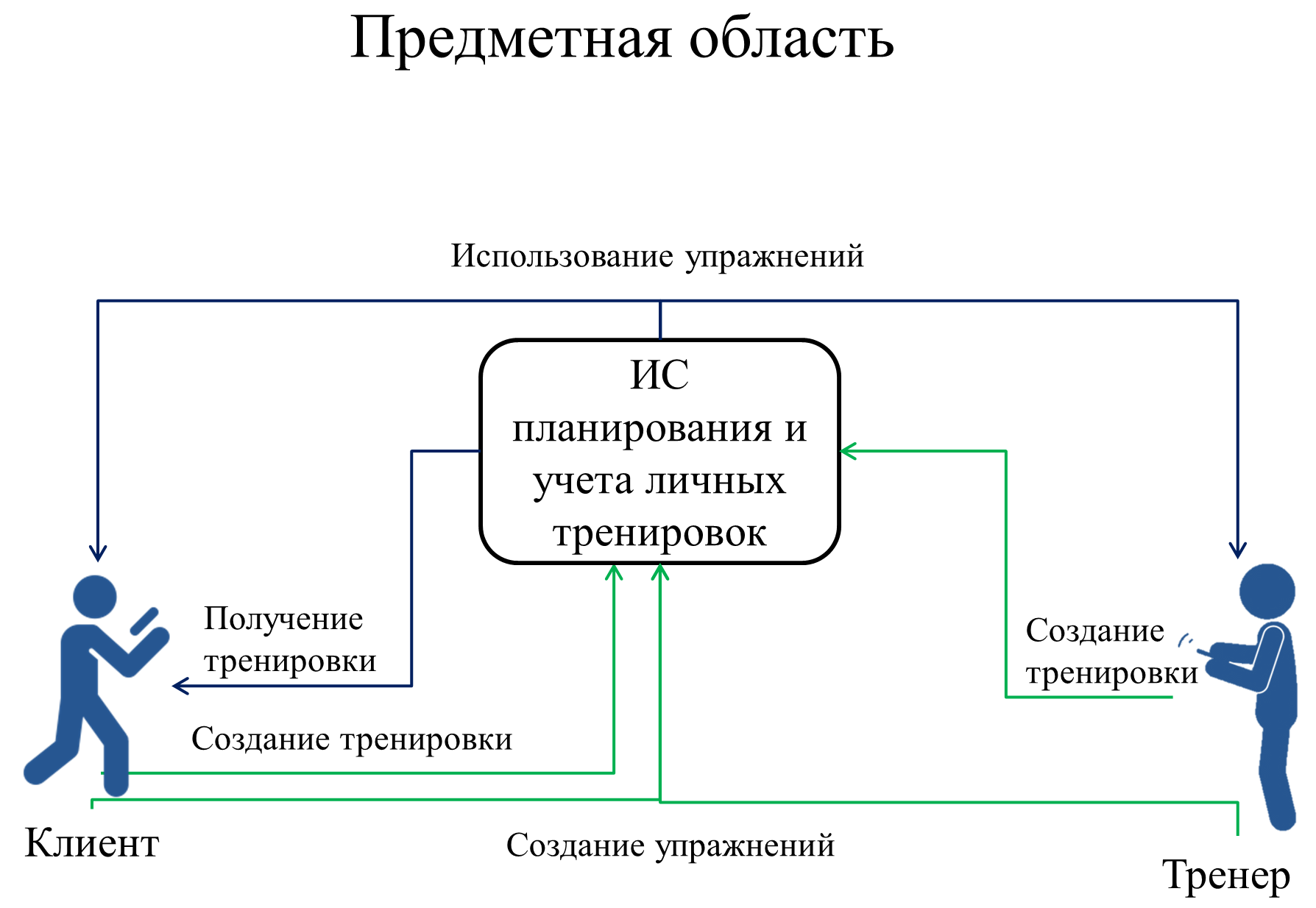
5. Инфологическая модель

6. Даталогическая модель

7. Пользовательские формы отображения

8. Пользовательские формы ввода





ПРИЛОЖЕНИЕ В   
ТЕХНИЧЕСКОЕ ЗАДАНИЕ

**Министерство образования и науки Российской Федерации**

**Федеральное государственное бюджетное образовательное учреждение**

**высшего профессионального образования**

**«Московский государственный технический университет имени Н.Э. Баумана»**

**(МГТУ им. Н.Э.Баумана)**

Утверждаю Согласовано

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_Черненький М.В.

«\_\_\_»\_\_\_\_\_\_\_\_\_\_2019 г. «\_\_\_»\_\_\_\_\_\_\_\_\_\_2019 г.

**«Информационная система планирования и учёта личных тренировок»**

техническое задание

(вид документа)

писчая бумага

(вид носителя)

9

(количество листов)

Исполнитель:

студент группы ИУ5-82

Фамилия И. О..

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

«\_\_\_» \_\_\_\_\_\_\_\_\_\_2019 г.

Москва, 2019 г.

**СОДЕРЖАНИЕ**

[1 Наименование 3](#_Toc199783522)

[2 Основание для разработки 3](#_Toc199783523)

[3 Исполнитель 3](#_Toc199783524)

[4 Назначение и цель разработки 3](#_Toc199783525)

[5 Содержание работы 3](#_Toc199783526)

[5.1 Задачи, подлежащие решению 3](#_Toc199783527)

[5.2 Требования к функциональности программного изделия 4](#_Toc199783528)

[5.3 Требования к архитектуре программного изделия 5](#_Toc199783529)

[5.4 Требования к составу программных компонентов 5](#_Toc199783530)

[5.5 Требования к базе данных 5](#_Toc199783531)

[5.6 Требования к входным и выходным данным 5](#_Toc199783532)

[5.6.1 Требования к входным данным 5](#_Toc199783533)

[5.6.2 Требования к выходным данным 6](#_Toc199783534)

[5.7 Требования к составу и характеристикам программных средств 6](#_Toc199783535)

[5.7.1 Требования к программному обеспечению сервера 6](#_Toc199783536)

[5.7.2 Требования к программному обеспечению клиента 6](#_Toc199783537)

[5.8 Требования к составу и характеристикам технических средств 6](#_Toc199783538)

[5.8.1. Требования к составу и характеристикам технических средств сервера 6](#_Toc199783539)

[5.8.2 Требования к составу и характеристикам технических средств клиента 7](#_Toc199783540)

[6 Этапы разработки 7](#_Toc199783541)

[7 Техническая документация, предъявляемая по окончании работы 7](#_Toc199783542)

[8 Порядок приемки работы 8](#_Toc199783543)

9 [Дополнительные условия 8](#_Toc199783544)

**Наименование**

Информационная система планирования и учета личных тренировок. Краткое название: ИСПУЛТ

1. **Основание для разработки**

Основанием для разработки является задание на ВКР, подписанное руководителем ВКР и утвержденное заведующим кафедрой ИУ5 МГТУ им. Н.Э. Баумана.

1. **Исполнитель**

Студент МГТУ им. Н.Э. Баумана группы ИУ5-72 Горбовцова Ксения Михайловна.

1. **Назначение и цель разработки**

Разрабатываемая система предназначена для оптимизации общения тренеров с их клиентами при удаленной работе, а также для составления тренировочных планов индивидуального использования.

1. **Содержание работы**
   1. **Задачи, подлежащие решению**

* исследование предметной области;
* разработка программного изделия;
* отладка программного изделия;
* разработка инфологической модели базы данных;
* разработка даталогической модели базы данных;
* разработка графа диалога пользователя;
* исследование особенностей функционирования продукта в сети.
  1. **Требования к функциональности программного изделия**

ИСПУТ должна удовлетворять следующим требованиям:

Для пользователя системы должны быть обеспечены следующие возможности:

* + 1. Регистрация в системе;

1. Первичная регистрация;
2. Вход в систему;
3. Выход из системы;
   * 1. Создание собственных тренировок:
4. Добавление пользователей к тренировке;
5. Ввод данных тренировки
6. Добавление упражнений к тренировке;
   * 1. Создание упражнений:
7. Прикрепление оборудования к упражнению;
8. Ввод данных упражнения
9. Прикрепление параметра к упражнению;
   * 1. Работа с тренировками:
10. Просмотр тренировок с текущей датой;
11. Просмотр легенды тренировок;
12. Удаление тренировок;
13. Просмотр упражнений, входящих в тренировку;
    * 1. Работа с упражнениями:
14. Доступ к банку упражнений информационной системы;
15. Просмотр детализированной информации об упражнении;
16. Удаление упражнения из тренировки;
17. Поиск упражнения по названию в банке системы;
    * 1. Работа с другими зарегистрированными пользователями:
18. Поиск пользователя по email адресу;
19. Добавление пользователя в друзья;
20. Удаление пользователя из друзей;
21. Просмотр профиля пользователя;
    * 1. Работа с личным кабинетом:
22. Изменение фотографии;
23. Изменение пароля;
24. Изменение фамилии, имени и email адреса;

ПРИЛОЖЕНИЕ В   
ПРОГРАММА И МЕТОДИКА ИСПЫТАНИЙ

**Министерство образования и науки Российской Федерации**

**Федеральное государственное бюджетное образовательное учреждение**

**высшего профессионального образования**

**«Московский государственный технический университет имени Н.Э. Баумана»**

**(МГТУ им. Н.Э.Баумана)**

Утверждаю Согласовано

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_ Черненький М.В.

«\_\_\_»\_\_\_\_\_\_\_\_\_\_2019 г. «\_\_\_»\_\_\_\_\_\_\_\_\_\_2019 г.

**«Информационная система планирования и учёта личных тренировок»**

программа и методика испытаний

(вид документа)

писчая бумага

(вид носителя)

10

(количество листов)

Исполнитель:

студент группы ИУ5-82

Фамилия И.О..

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

«\_\_\_»\_\_\_\_\_\_\_\_\_\_2019 г.

Москва ,2019 г.