

Федеральное государственное автономное образовательное
учреждение высшего образования
«Национальный исследовательский университет ИТМО»

Факультет программной инженерии и компьютерной техники

**Техническое задание на написание выпускной
квалификационной работы**

Работу выполнил студент группы Р3415:
Барсуков Максим Андреевич

Преподаватель:
Маркина Татьяна Анатольевна

г. Санкт-Петербург
2026 г.

1. Наименование

Полное наименование: Выпускная квалификационная работа на тему «Проектирование и разработка распределенного хранилища данных для семантической сети».

Краткое наименование: ВКР «Проектирование и разработка распределенного хранилища данных для семантической сети».

Условное обозначение: ВКР-DSNS-2026.

2. Назначение

Главная (административная) цель проекта — системное планирование, выполнение и успешная защита выпускной квалификационной работы для завершения обучения по программе бакалавриата и получения квалификации.

Данный проект (ВКР) предназначен для планирования, определения целей, задач, содержания, этапов и порядка выполнения выпускной квалификационной работы, выполнения и контроля всех этапов создания выпускной квалификационной работы, итогом которой является:

1. Исследование архитектурных подходов и моделей данных для распределенных систем, работающих со структурированной семантической информацией (RDF, графовые базы, онтологии).
2. Разработанное и функционирующее программное обеспечение — прототип распределенного хранилища данных, способного эффективно хранить, индексировать и выполнять базовые запросы к семантическим данным, поддерживающего репликацию и шардирование. Результат работы должен продемонстрировать применимость выбранных решений.

Результат работы должен продемонстрировать обоснованность выбранных проектных решений и соответствие требованиям, предъявляемым к выпускной квалификационной работе.

Область применения проекта: учебно-исследовательская деятельность в области распределенных систем, семантических сетей и управления большими объемами структурированных данных.

3. Основание для разработки

Документами, на основании которых производится разработка, служат:

1. Требования к выпускным квалификационным работам [Электронный ресурс] – URL: <http://student.itmo.ru/files/1314>
2. ГОСТ 7.32-2017. Межгосударственный стандарт. Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления
3. Principles of Semantic Networks: Explorations in the Representation of Knowledge, John F. Sowa (1991)

4. Функции проекта

Проект по написанию ВКР должен обеспечить выполнение следующих ключевых функций:

1. Аналитическая функция: Проведение обзора современного состояния проблематики (распределенные хранилища, семантические сети, графовые СУБД, протоколы обмена).
2. Проектно-технологическая функция: Формулирование требований, проектирование архитектуры хранилища (выбор стека технологий, модели данных, алгоритмов репликации/шардирования, протокола узлов).
3. Реализационная функция: Написание исходного кода ключевых компонентов системы (модуль хранения RDF-триплетов, модуль кластеризации/коммуникации, простой query-интерфейс).
4. Верификационная функция: Тестирование прототипа на корректность, производительность и устойчивость к сбоям. Сравнение с базовыми решениями.
5. Документационная функция: Оформление пояснительной записи, отражающей все этапы работы, принятые решения и результаты.
6. Презентационная функция: Подготовка материалов для защиты (презентация, доклад).

В рамках проекта должен быть разработан прототип ПО, реализующий следующие базовые функции:

1. Функция хранения и управления семантическими данными: прием и сохранение RDF-триплетов (графовых данных) в устойчивом формате, базовая индексация данных для ускорения поиска по субъекту, предикату и объекту (SPO-индексы), удаление и обновление записей.

2. Функция распределенного выполнения запросов: обработка простых запросов к графу (например, шаблонных запросов вида «Субъект-Предикат-?Объект») в рамках одного узла, поддержка механизма маршрутизации и выполнения составных запросов, требующих обращения к данным на нескольких узлах кластера.
3. Функция обеспечения отказоустойчивости и согласованности: репликация данных между узлами кластера для повышения доступности, реализация базового протокола согласования состояния (например, на основе лидера или gossip-протокола) для обеспечения целостности данных при записи, обнаружение отказов узлов и перераспределение их нагрузки.
4. Функция кластеризации и масштабирования: динамическое присоединение новых узлов к кластеру, автоматическое распределение данных между узлами (шардирование) на основе выбранной стратегии (например, по хэшу ключа), предоставление API для мониторинга состояния кластера (состав узлов, метрики нагрузки).
5. Функция предоставления внешнего доступа (API): предоставление сетевого интерфейса для основных операций: вставка данных (INSERT), выборка (SELECT), проверка состояния системы (HEALTH).

5. Структура проекта

Проект включает три взаимосвязанных компонента:

1. Документальная часть:
 - Пояснительная записка ВКР.
 - Презентация для защиты.
 - Текст доклада.
2. Практическая/исследовательская часть (артефакт):
 - Исходный код прототипа распределенного хранилища.
 - Конфигурационные файлы и скрипты развертывания (Docker Compose).
 - Набор тестовых данных (семантические графы) и скрипты для нагружочного тестирования.
3. Процессуальная часть:
 - План-график.
 - Материалы этапов согласования с руководителем.
 - Результаты предзащиты.

Обязательная структура текстовой части ВКР:

- титульный лист
- задание
- аннотация
- содержание
- текст ВКР
 - введение
 - основная часть
 - заключение
- список использованных источников

6. Пользовательский интерфейс

Прямого графического интерфейса у проекта «Написание ВКР» нет. Его основными «интерфейсами» являются структура и оформление итоговых документов (пояснительной записки, презентации), обеспечивающие удобство восприятия. Отсутствие требований к пользовательскому интерфейсу для текстовой части ВКР обусловлено тем, что текстовая часть оформляется в соответствии с установленными шаблонами и требованиями, не подразумевающими разработку специального интерфейса.

Однако, разрабатываемый в рамках проекта программный продукт (хранилище) будет предоставлять:

- Программный интерфейс (API): RESTful API или gRPC-интерфейс для выполнения операций вставки и выборки данных (SPARQL-подобные запросы или упрощенные запросы к графу).
- Интерфейс командной строки (CLI): Утилиты для администрирования кластера, запуска узлов и выполнения базовых операций.

Требования к этим интерфейсам будут детализированы во время написания ТЗ разрабатываемого ПО.

7. Требования к надежности, безопасности и условия эксплуатации

Надежность проекта:

1. Проект должен быть выполнен в установленные сроки.
2. Пояснительная записка и код должны сохраняться в нескольких экземплярах (локально, в облаке, в системе контроля версий Git – GitHub/GitLab).
3. Надежность разрабатываемого ПО: прототип должен обеспечивать сохранность данных при штатном завершении работы и иметь механизмы восстановления после сбоев отдельных узлов (заложено в архитектуре).

Безопасность:

1. Академическая безопасность: обязательная проверка текста пояснительной записи в системе антиплагиата. Оформление заимствований в соответствии с правилами цитирования.
2. Безопасность ПО: в прототипе на этапе ВКР требования к криптографической защите каналов передачи данных могут быть опущены для упрощения, но архитектура должна допускать их добавление. Достаточно базовой аутентификации узлов в кластере.

Условия эксплуатации:

1. Аппаратные: Разработка ведется на персональном компьютере студента с процессором не менее 4 ядер, 16 ГБ ОЗУ, 100 ГБ свободного места на SSD. Тестирование кластера предполагает развертывание на нескольких виртуальных машинах или в контейнерах.
2. Программные: ОС – Linux (Ubuntu 22.04+). Стек разработки: язык программирования [Java 17+], СУБД RocksDB (используется как бэкенд), Docker, Git, IDE (VS Code, IntelliJ IDEA), средства документирования (Markdown, Word).

8. Документация

В рамках проекта должна быть разработана следующая документация:

1. Полный текст работы с обоснованием принятых решений.
2. Презентация: 12-15 слайдов, текст доклада (3-5 страниц).
3. Рецензия научного руководителя: Оценка выполненной работы.
4. Пояснительная записка к ВКР. Оформляется в соответствии с ГОСТ 7.32-2017 и методическими указаниями кафедры. Объем: 60-80 страниц.
5. Техническая документация: Архитектура, API, руководство по разработке. Документация должна быть разработана в соответствии с ГОСТ Р ИСО/МЭК 15910-2002 ПРОЦЕСС СОЗДАНИЯ ДОКУМЕНТАЦИИ ПОЛЬЗОВАТЕЛЯ ПРОГРАММНОГО СРЕДСТВА.
6. Внутренняя документация к программному коду: README.md с инструкцией по сборке, настройке и запуску, описание архитектуры (ADR – Architecture Decision Record) в виде Markdown-файлов, комментарии в коде на английском или русском языке в формате, поддерживаемом генераторами документации (например, GoDoc, Javadoc, Sphinx).

9. Стадии и этапы разработки

Этап 1: Подготовительный.

1. Выбор научного руководителя, предварительное определение темы. Срок выполнения: октябрь 2025. Результат: научный руководитель закреплен.
2. Обсуждение темы ВКР с научным руководителем. Сбор литературы, уточнение формулировки темы. Срок выполнения: ноябрь-декабрь 2025. Результат: список источников, обоснование темы.
3. Утверждение темы ВКР. Срок выполнения: январь 2026. Результат: тема утверждена научным руководителем, секретарем ГЭК.

Этап 2: Проектно-аналитический.

1. Разработка и согласование настоящего ТЗ. Срок выполнения: февраль 2026. Результат: данный документ.
2. Глубокий анализ предметной области и аналогов. Срок выполнения: февраль-март 2026. Результат: раздел ВКР «Аналитический обзор».
3. Формулирование требований и проектирование архитектуры системы. Срок выполнения: февраль-март 2026. Результат: спецификация требований, схемы архитектуры.

Этап 3: Реализация.

1. Написание кода базовых модулей (хранилище, сеть). Срок выполнения: март-апрель 2026. Результат: работающий каркас (single-node).
2. Реализация механизмов распределённости. Срок выполнения: март-апрель 2026. Результат: прототип, работающий в кластере.

Этап 4: Тестирование и оформление.

1. Функциональное и нагрузочное тестирование. Срок выполнения: апрель 2026. Результат: отчет о тестировании, графики.
2. Написание основной части пояснительной записки. Срок выполнения: апрель 2026. Результат: черновик пояснительной записки.
3. Написание документации к программному коду. Результат: README.md, описание API, комментарии в коде.

Этап 5: Документирование, подготовка к защите, защита.

1. Подготовка презентации и текста доклада. Срок выполнения: апрель-май 2026. Результат: Черновик презентации (12-15 слайдов) и доклада.
2. Предзащита, доработка по замечаниям. Результат: допуск к защите.

3. Окончательное оформление, проверка на антиплагиат, подготовка к защите. Результат: готовый пакет документов для ГЭК.
4. Защита ВКР. Результат: положительная оценка ГЭК.

10. Порядок контроля и приёма

После каждого этапа разработки осуществляется анализ и, если возможно, тестирование полученных решений. Соответствующие выводы должны содержаться в ВКР.

1. Текущий контроль: осуществляется студентом и научным руководителем в ходе регулярных консультаций. Отчеты о прогрессе разработки.
2. Промежуточный контроль: после утверждения данного ТЗ контроль будет осуществляться в соответствии с графиком. Сдача на проверку разделов пояснительной записи и демонстрация работающих модулей ПО. Показ работающих компонентов.
3. Предварительная приемка: выступление на предзащите. Получение допуска к защите.
4. Итоговая приемка (защита): проверка ВКР на антиплагиат и защита работы перед Государственной экзаменацационной комиссией (ГЭК). Критерием успешной приемки является положительная оценка комиссии, отраженная в протоколе заседания ГЭК.