

Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт

Курсовая работа
«Генерация текста НИР с помощью ИИ»

**Тема работы: «Сравнительный анализ методов автоматического выделения
ключевых сущностей и связей для построения тематических графов знаний
в научных областях»**

по дисциплине «Автоматизация научных исследований»

Выполнил
студент гр. № 5040102/50201

Худина А.А.

Преподаватель: Новиков Ф.А.

Санкт-Петербург
2025 г.

Аннотация

Статья посвящена проектированию архитектуры и разработке прототипа веб-инструмента KnowledgeGraph Vis для визуализации тематических графов знаний, построенных на основе научных публикаций. Цель работы — формализация требований и создание проектной документации для системы автоматического анализа исследовательских ландшафтов. В ходе работы применялись методы объектно-ориентированного проектирования (UML), адаптивной веб-разработки и визуальной аналитики. В результате разработаны диаграммы вариантов использования и классов системы, создан функциональный прототип презентационного веб-сайта и сформирован методологический базис для последующей реализации алгоритмической части. Практическая значимость работы заключается в предложении структурированного подхода к созданию инструментов научометрического анализа для исследователей и аналитиков данных.

Abstract

The article focuses on the architecture design and prototyping of the KnowledgeGraph Vis web tool for visualizing thematic knowledge graphs built from scientific publications. The aim of the work is to formalize requirements and create design documentation for a system of automatic research landscape analysis. The work applies methods of object-oriented design (UML), adaptive web development, and visual analytics. As a result, use case and class diagrams of the system have been developed, a functional prototype of a promotional website has been created, and a methodological basis for the subsequent implementation of algorithmic components has been formed. The practical significance of the work lies in proposing a structured approach to creating scientometric analysis tools for researchers and data analysts.

Введение

Современный этап развития науки характеризуется экспоненциальным ростом объема публикуемых материалов. Это создает существенные сложности для исследователей, стремящихся выявить ключевые тенденции, центральные работы и скрытые взаимосвязи внутри быстроразвивающихся дисциплин. В данном контексте актуальной задачей становится разработка специализированных инструментов, способных автоматически структурировать большие корпуса научных текстов и предоставлять результаты анализа в наглядной, интерактивной форме.

Одним из перспективных подходов к решению этой задачи является построение и визуализация тематических графов знаний. В таких графах узлы представляют сущности различного типа (концепции, методы, авторы, публикации), а ребра отражают семантические или цитационные связи между ними. Подобное представление позволяет перейти от анализа отдельных документов к исследованию структуры целых научных направлений.

Данная работа посвящена проектированию архитектуры и разработке прототипа веб-инструмента KnowledgeGraph Vis, целью которого является визуализация исследовательских ландшафтов на основе графов знаний. В отличие от существующих решений, акцент делается на создании универсальной, гибкой системы с четко formalизованной структурой, что обеспечивает основу для последующей интеграции различных алгоритмов анализа.

1 Обзор существующих решений и постановка задачи

Анализ предметной области выявил, что, несмотря на наличие ряда инструментов для сетевого анализа (таких как Gephi [8]) и научометрических платформ, существует потребность в специализированных решениях, сочетающих автоматическое построение графов из текстовых данных с интерактивной веб-визуализацией. Многие существующие подходы либо требуют значительной ручной подготовки данных, либо ограничены конкретными типами связей (например, только цитирования [6]).

Таким образом, была сформулирована задача проектирования системы, которая:

1. Позволяет загружать корпуса научных публикаций в стандартных форматах.
2. Обеспечивает возможность интеграции различных алгоритмов обработки естественного языка (NLP) и сетевого анализа для выделения сущностей и связей [3, 7].
3. Предоставляет пользователю интерактивный интерфейс для визуального исследования построенного графа, его фильтрации и анализа основных метрик [4, 9].
4. Имеет модульную архитектуру, позволяющую расширять функциональность.

Объектом исследования является процесс проектирования веб-инструментов для визуальной аналитики научных данных. Предмет исследования — архитектура и пользовательский интерфейс системы KnowledgeGraph Vis.

Для достижения поставленной цели были определены следующие задачи:

1. Провести анализ требований к системе и разработать диаграмму вариантов использования.
2. Спроектировать логическую структуру системы и разработать диаграмму классов.
3. Реализовать функциональный прототип презентационного веб-интерфейса (лендинга) системы.
4. Сформировать методологическую базу для будущей реализации алгоритмических компонентов.

2 Методы и материалы

В работе использовался комплекс методов, характерных для проектирования информационных систем:

1. **Моделирование требований и бизнес-процессов:** Для формализации функциональных требований и взаимодействия пользователей с системой была применена нотация UML Use Case Diagram.
2. **Объектно-ориентированное проектирование:** Для описания статической структуры данных и компонентов системы была использована нотация UML Class Diagram.
3. **Адаптивная веб-разработка:** Для создания прототипа пользовательского интерфейса применялись технологии HTML5, CSS3 и JavaScript (ES6+). Для обеспечения адаптивности и ускорения разработки использовался фреймворк Bootstrap 5.2.
4. **Методология визуальной аналитики:** При проектировании интерфейса были учтены принципы интерактивного исследования данных, включая возможность динамической фильтрации, масштабирования и анализа метрик графа [2].

Материалом для проектирования послужили требования к проекту KnowledgeGraph Vis, а также существующие практики в области наукометрии [2], NLP [3, 7] и визуализации графов [8].

3 Результаты и их обсуждение

3.1 Диаграмма вариантов использования системы

Для спецификации функциональных требований была разработана диаграмма вариантов использования. Основными акторами системы были идентифицированы: Исследователь/Аналитик (ключевой пользователь), Администратор системы и Внешняя система (источник данных).

Диаграмма определяет такие ключевые сценарии, как загрузка корпуса публикаций, настройка параметров анализа, автоматическое построение и интерактивная визуализация графа, анализ сетевых метрик и экспорт результатов. Данная модель явно разделяет роли и уточняет границы системы.

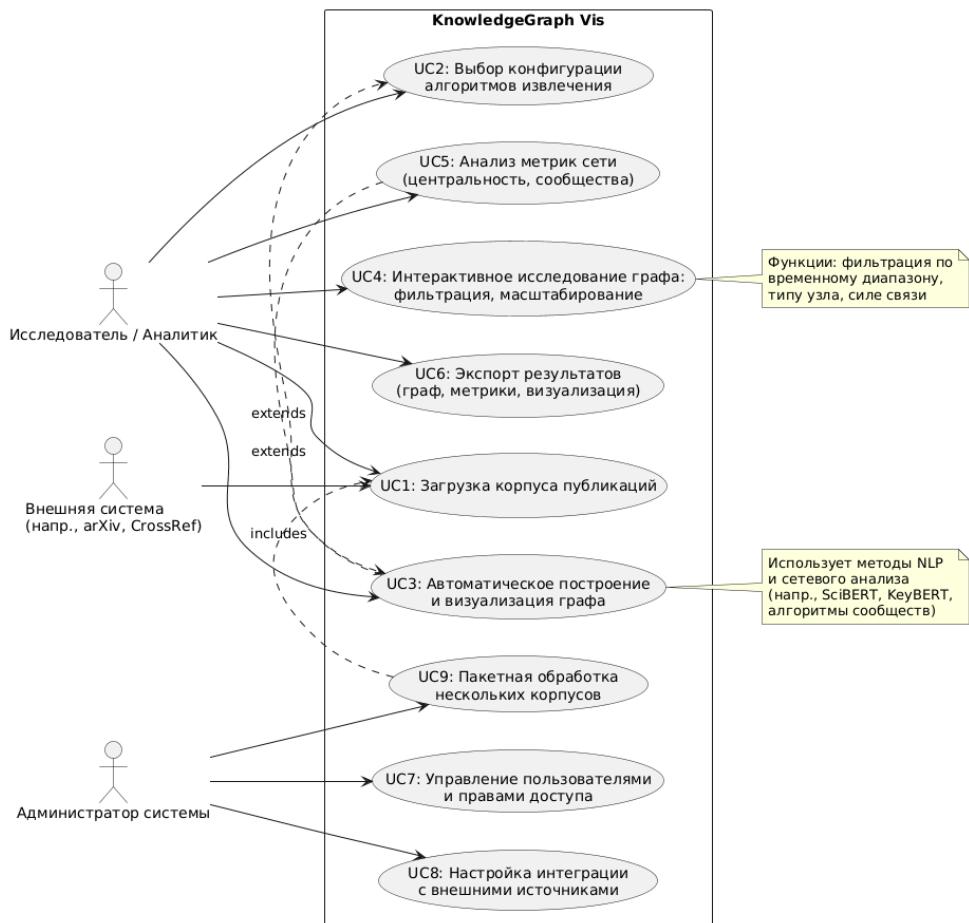


Рис. 1. Диаграмма вариантов использования системы KnowledgeGraph Vis.

3.2 Диаграмма классов системы

На основе анализа предметной области была разработана диаграмма классов, формализующая ключевые сущности и их взаимосвязи. Центральными концепциями стали:

- **Данные:** Классы Publication, Author, Corpus.
- **Граф знаний:** Классы KnowledgeGraph, GraphNode (с атрибутом `type: NodeType`), GraphEdge (с атрибутом `type: EdgeType`).
- **Обработка:** Классы ExtractionAlgorithm (абстракция для алгоритмов NLP и сетевого анализа) и AnalysisConfiguration.
- **Пользователи:** Классы Researcher и Administrator, наследующиеся от User.

Диаграмма фиксирует отношения композиции между графом и его элементами, а также ассоциации, связывающие узлы графа с исходными сущностями. Эта модель служит основой для будущей объектно-ориентированной реализации.

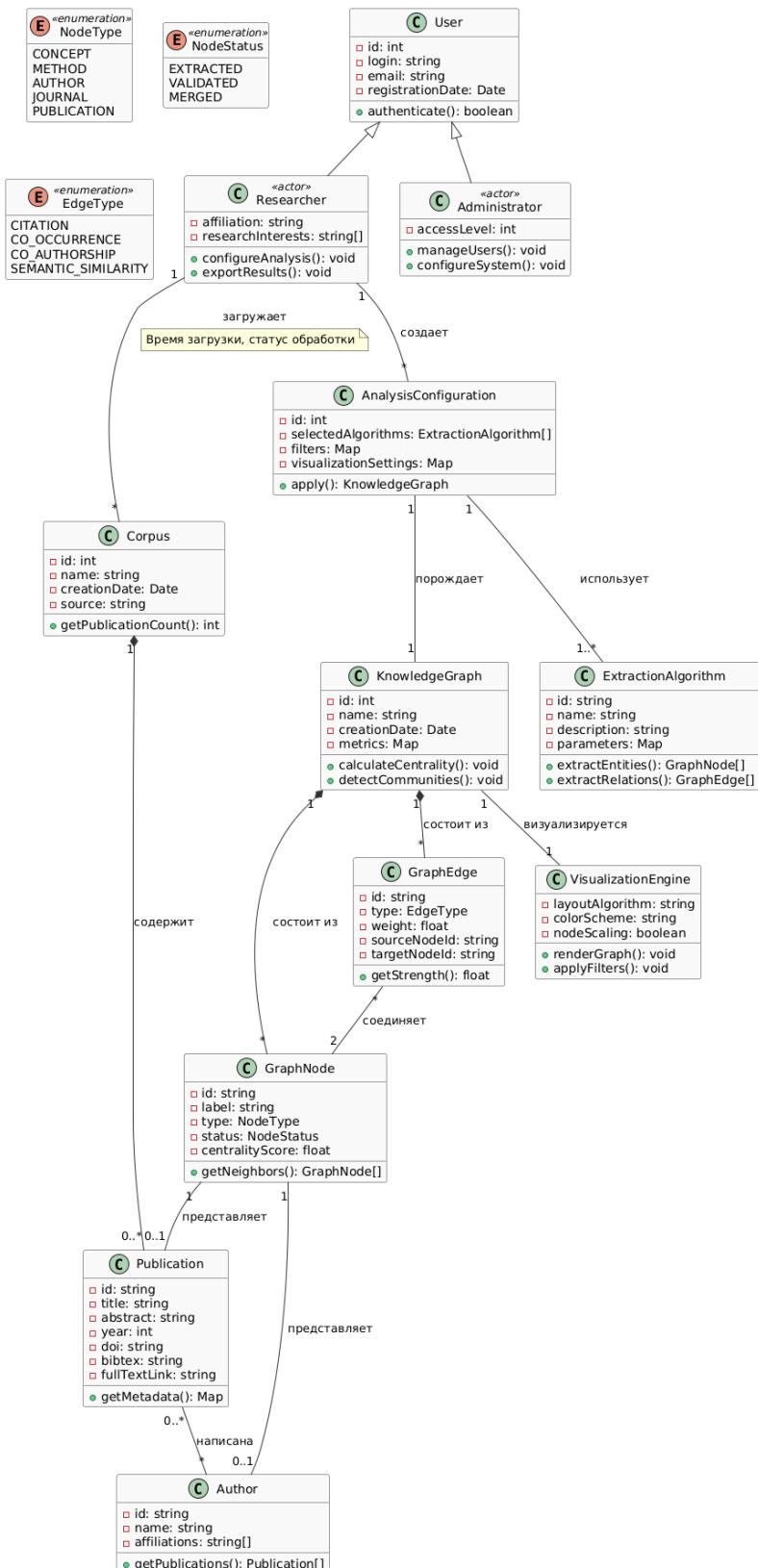


Рис. 2. Диаграмма классов системы KnowledgeGraph Vis.

3.3 Прототип презентационного веб-интерфейса

В рамках практической реализации был разработан полнофункциональный, адаптивный лендинг проекта. Сайт реализован с использованием семантической HTML-разметки, CSS-переменных для управления цветовой палитрой и ванильного JavaScript для интерактивности. Лендинг включает следующие блоки: навигация, герой-секция, описание возможностей, стилизованное демо-окно с имитацией графа, описание шагов работы, форма сбора обратной связи и подвал. Функциональность

включает плавную прокрутку и валидацию формы. Полный исходный код представлен в Приложении А.

Кодовая база демонстрирует реализацию ключевых принципов, заложенных в диаграммах: модульность, интерактивность и ориентацию на целевую аудиторию. Созданный прототип является работающим артефактом, подтверждающим жизнеспособность предложенной архитектуры пользовательского интерфейса.

3.4 Методологическая база для алгоритмической реализации

Для обеспечения перехода к следующему этапу — реализации и сравнению алгоритмов извлечения сущностей и связей — был сформирован структурированный набор поисковых запросов для академического репозитория arXiv. Запросы охватывают ключевые тематические фасеты: построение научных графов знаний, извлечение сущностей и отношений, анализ и кластеризация сетей, визуализация графов, наукометрия. Этот набор обеспечивает систематический доступ к современным исследованиям, необходимым для обоснованного выбора и реализации алгоритмов в разрабатываемой системе.

Заключение

В результате проведенной работы была спроектирована архитектура веб-инструмента KnowledgeGraph Vis для визуализации исследовательских ландшафтов. Выполнены следующие задачи:

1. Проанализированы требования к системе и формализованы сценарии ее использования с помощью UML диаграммы вариантов использования.
2. Спроектирована статическая структура системы и данных с использованием UML диаграммы классов, что определило ключевые сущности и их взаимосвязи.
3. Разработан и реализован адаптивный функциональный прототип презентационного веб-интерфейса системы, подтверждающий основные принципы взаимодействия с пользователем.
4. Сформирован методологический базис для последующего этапа работ — комплексный набор поисковых запросов для изучения современных алгоритмических решений.

Полученные результаты — проектная документация и работающий прототип — создают прочный фундамент для дальнейшей разработки полнофункциональной системы. Последовательная реализация следующих этапов, начиная с интеграции алгоритмов NLP и сетевого анализа на основе проведенного литературного обзора, позволит реализовать инструмент, способный эффективно помогать исследователям в анализе динамики научных направлений.

Список использованных источников

1. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. Translating Embeddings for Modeling Multi-relational Data / A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, O. Yakhnenko // Advances in Neural Information Processing Systems 26 (NIPS 2013). – 2013. – P. 2787–2795.
2. Chen, C. Science Mapping: A Systematic Review of the Literature / C. Chen // Journal of Data and Information Science. – 2017. – Vol. 2, № 2. – P. 1–40. DOI: 10.1515/jdis-2017-0006.
3. Mikolov, T., Chen, K., Corrado, G., & Dean, J. Efficient Estimation of Word Representations in Vector Space / T. Mikolov, K. Chen, G. Corrado, J. Dean // arXiv preprint. – 2013. – arXiv:1301.3781.
4. Newman, M. E. J. Modularity and community structure in networks / M. E. J. Newman // Proceedings of the National Academy of Sciences. – 2006. – Vol. 103, № 23. – P. 8577–8582. DOI: 10.1073/pnas.0601602103.
5. Perozzi, B., Al-Rfou, R., & Skiena, S. DeepWalk: Online Learning of Social Representations / B. Perozzi, R. Al-Rfou, S. Skiena // Proceedings of the 20th ACM SIGKDD International Conference

- on Knowledge Discovery and Data Mining (KDD '14). – New York, NY, USA : ACM, 2014. – P. 701–710. DOI: 10.1145/2623330.2623732.
6. Small, H. Co-citation in the scientific literature: A new measure of the relationship between two documents / H. Small // Journal of the American Society for Information Science. – 1973. – Vol. 24, № 4. – P. 265–269. DOI: 10.1002/asi.4630240406.
 7. Luan, Y., He, L., Ostendorf, M., & Hajishirzi, H. Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction / Y. Luan, L. He, M. Ostendorf, H. Hajishirzi // arXiv preprint. – 2018. – arXiv:1808.09602.
 8. Bastian, M., Heymann, S., & Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks / M. Bastian, S. Heymann, M. Jacomy // Proceedings of the International AAAI Conference on Web and Social Media. – 2009. – Vol. 3, № 1. – P. 361–362. DOI: 10.1609/icwsm.v3i1.13937.
 9. Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. Fast unfolding of communities in large networks / V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre // Journal of Statistical Mechanics: Theory and Experiment. – 2008. – Vol. 2008, № 10. – P. P10008. DOI: 10.1088/1742-5468/2008/10/P10008.
 10. Shi, F., Chen, L., Han, J., & Childs, P. A data-driven text mining and semantic network analysis for design information retrieval / F. Shi, L. Chen, J. Han, P. Childs // Journal of Mechanical Design. – 2017. – Vol. 139, № 11. – P. 111402. DOI: 10.1115/1.4037649.

Приложение А

Исходный код презентационного веб-сайта (лендинга) проекта KnowledgeGraph Vis:

```
<!DOCTYPE html>
<html lang="ru">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>KnowledgeGraph Vis | Визуализация исследовательских ландшафтов</title>
    <!-- Bootstrap 5.2 CDN -->
    <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.2.3/dist/css/bootstrap.min.css" rel="stylesheet">
        <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap-icons@1.9.1/font/bootstrap-icons.css">
    <style>
        :root {
            --color-primary: #0d1b2a;
            --color-secondary: #e0e1dd;
            --color-accent: #2ec4b6;
            --color-accent-light: #cbf3f0;
            --color-white: #ffffff;
        }
        body {
            font-family: 'Segoe UI', system-ui, -apple-system, sans-serif;
            color: var(--color-primary);
            line-height: 1.6;
        }
        .navbar-brand {
            font-weight: 700;
            color: var(--color-accent) !important;
        }
        .hero-section {
            background: linear-gradient(135deg, var(--color-primary) 0%, #1b2b3a 100%);
            color: var(--color-secondary);
            padding: 6rem 0;
        }
        .btn-primary {
            background-color: var(--color-accent);
            border-color: var(--color-accent);
            padding: 0.75rem 2rem;
        }
    </style>

```

```

        font-weight: 600;
    }
    .btn-primary:hover {
        background-color: #26a89a;
        border-color: #26a89a;
    }
    .feature-card {
        border: none;
        border-radius: 0.5rem;
        transition: transform 0.3s ease;
        height: 100%;
        background: var(--color-white);
        box-shadow: 0 4px 12px rgba(13, 27, 42, 0.08);
    }
    .feature-card:hover {
        transform: translateY(-5px);
    }
    .feature-icon {
        font-size: 2.5rem;
        color: var(--color-accent);
        margin-bottom: 1rem;
    }
    .demo-container {
        background: var(--color-primary);
        border-radius: 0.75rem;
        padding: 2rem;
        position: relative;
        overflow: hidden;
    }
    .demo-window {
        background: #1a1a2e;
        border-radius: 0.5rem;
        height: 400px;
        position: relative;
        overflow: hidden;
    }
    .graph-node {
        position: absolute;
        border-radius: 50%;
        background: var(--color-accent);
        animation: pulse 2s infinite;
    }
    @keyframes pulse {
        0% { opacity: 0.7; }
        50% { opacity: 1; }
        100% { opacity: 0.7; }
    }
    .step-timeline {
        position: relative;
        padding-left: 2rem;
    }
    .step-timeline::before {
        content: '';
        position: absolute;
        left: 0;
        top: 0;
        bottom: 0;
        width: 3px;
        background: var(--color-accent-light);
    }
    .step-item {
        position: relative;
        margin-bottom: 2rem;
    }
    .step-item::before {
        content: '';

```

```

        position: absolute;
        left: -2.3rem;
        top: 0.5rem;
        width: 1rem;
        height: 1rem;
        border-radius: 50%;
        background: var(--color-accent);
        border: 3px solid var(--color-white);
        box-shadow: 0 0 0 3px var(--color-accent-light);
    }
    .footer {
        background: var(--color-primary);
        color: var(--color-secondary);
        padding: 3rem 0;
    }
    .form-control:focus {
        border-color: var(--color-accent);
        box-shadow: 0 0 0 0.25rem rgba(46, 196, 182, 0.25);
    }
    @media (max-width: 768px) {
        .hero-section {
            padding: 4rem 0;
        }
        .step-timeline::before {
            left: 1.5rem;
        }
        .step-item::before {
            left: -0.8rem;
        }
    }

```

</style>

</head>

<body>

<!-- Header -->

<nav class="navbar navbar-expand-lg navbar-light bg-white sticky-top shadow-sm">

<div class="container">

KG Vis

<button class="navbar-toggler" type="button" data-bs-toggle="collapse" data-bs-target="#navbarNav">

</button>

<div class="collapse navbar-collapse" id="navbarNav">

<ul class="navbar-nav ms-auto">

<li class="nav-item">Возможности

<li class="nav-item">Демо

<li class="nav-item">Как это

работает

<li class="nav-item">О

проекте

<li class="nav-item">Контакты

</div>

</div>

</nav>

<!-- Hero Section -->

<section class="hero-section">

<div class="container">

<div class="row align-items-center">

<div class="col-lg-6">

<h1 class="display-4 fw-bold mb-4">KnowledgeGraph Vis
<span

class="text-accent">Визуализируйте структуру науки.</h1>

<p class="lead mb-4">Иновационный веб-инструмент для анализа

научных публикаций. Преобразуйте коллекции статей в интерактивные графы знаний,

выявляйте ключевых авторов, центральные концепции и скрытые исследовательские тренды.</p>
 Попробовать демо
 Узнать больше

```

        </div>
        <div class="col-lg-6">
            <div class="demo-window mt-5 mt-lg-0" id="hero-graph">
                <!-- Граф будет сгенерирован JavaScript -->
            </div>
        </div>
    </div>
</section>


<section id="features" class="py-5 bg-light">
    <div class="container">
        <h2 class="text-center mb-5 fw-bold">Ключевые возможности</h2>
        <div class="row g-4">
            <div class="col-md-6 col-lg-3">
                <div class="feature-card p-4">
                    <div class="feature-icon"><i class="bi bi-diagram-3"></i></div>
                    <h4>Интерактивная визуализация</h4>
                    <p>Динамическое исследование графа с масштабированием, фильтрацией по времени и типу узла.</p>
                </div>
            </div>
            <div class="col-md-6 col-lg-3">
                <div class="feature-card p-4">
                    <div class="feature-icon"><i class="bi bi-graph-up"></i></div>
                    <h4>Сетевой анализ</h4>
                    <p>Расчет метрик центральности, выявление сообществ и структурных паттернов.</p>
                </div>
            </div>
            <div class="col-md-6 col-lg-3">
                <div class="feature-card p-4">
                    <div class="feature-icon"><i class="bi bi-file-earmark-arrow-up"></i></div>
                    <h4>Гибкая загрузка данных</h4>
                    <p>Поддержка форматов BibTeX, CSV и интеграция с репозиториями arXiv, Crossref.</p>
                </div>
            </div>
            <div class="col-md-6 col-lg-3">
                <div class="feature-card p-4">
                    <div class="feature-icon"><i class="bi bi-download"></i></div>
                    <h4>Экспорт результатов</h4>
                    <p>Сохранение графов, метрик и визуализаций для дальнейшего анализа и отчетности.</p>
                </div>
            </div>
        </div>
    </div>
</section>


<section id="demo" class="py-5">
    <div class="container">
        <h2 class="text-center mb-5 fw-bold">Посмотрите, как это работает</h2>
        <div class="demo-container">
            <div class="demo-window" id="demo-graph">
                <!-- Граф будет сгенерирован JavaScript -->
            </div>
        </div>
    </div>
</section>
```

```

<p class="text-center text-white mt-3">Интерактивная демо-версия
появится в ближайшее время. Подпишитесь, чтобы узнать первым.</p>
</div>
</div>
</section>

<!-- Steps -->
<section id="steps" class="py-5 bg-light">
<div class="container">
    <h2 class="text-center mb-5 fw-bold">Как это работает</h2>
    <div class="step-timeline">
        <div class="step-item">
            <h4>1. Загрузите данные</h4>
            <p>Загрузите корпус научных публикаций в формате BibTeX или CSV.
Система автоматически извлечет метаданные.</p>
        </div>
        <div class="step-item">
            <h4>2. Настройте параметры анализа</h4>
            <p>Выберите алгоритмы для выделения сущностей и связей, настройте
параметры фильтрации и визуализации.</p>
        </div>
        <div class="step-item">
            <h4>3. Исследуйте и визуализируйте</h4>
            <p>Анализируйте интерактивный граф, выявляйте кластеры, ключевые
работы и тренды в вашей области.</p>
        </div>
    </div>
</div>
</section>

<!-- CTA Form -->
<section id="contact" class="py-5">
<div class="container">
    <div class="row justify-content-center">
        <div class="col-lg-8 text-center">
            <h2 class="fw-bold mb-4">Готовы исследовать по-новому?</h2>
            <p class="mb-5">Оставьте email, чтобы получить уведомление о
запуске бета-теста и подробный дайджест возможностей KnowledgeGraph Vis.</p>
            <form id="subscribe-form" class="row g-3 justify-content-center">
                <div class="col-auto">
                    <input type="email" class="form-control form-control-lg"
id="email-input" placeholder="Ваш email" required>
                </div>
                <div class="col-auto">
                    <button type="submit" class="btn btn-primary btn-lg">Подписаться</button>
                </div>
            </form>
            <p class="small text-muted mt-3">Мы уважаем вашу
конфиденциальность. Ваш email будет использован только для уведомлений о проекте.</p>
        </div>
    </div>
</div>
</section>

<!-- Footer -->
<footer class="footer">
<div class="container">
    <div class="row">
        <div class="col-lg-4 mb-4 mb-lg-0">
            <h3 class="mb-3">KG Vis</h3>
            <p>Инструмент для визуализации и анализа исследовательских
ландшафтов на основе графов знаний.</p>
        </div>
        <div class="col-lg-4 mb-4 mb-lg-0">
            <h5 class="mb-3">Ссылки</h5>

```

```

        <a href="#features" class="d-block text-secondary mb-
2">Возможности</a>
            <a href="#demo" class="d-block text-secondary mb-2">Демо</a>
            <a href="#contact" class="d-block text-secondary mb-2">Контакты</a>
        </div>
        <div class="col-lg-4">
            <h5 class="mb-3">Социальные сети</h5>
            <a href="#" class="text-secondary me-3"><i class="bi bi-twitter fs-
5"></i></a>
            <a href="#" class="text-secondary me-3"><i class="bi bi-github fs-
5"></i></a>
            <a href="#" class="text-secondary me-3"><i class="bi bi-telegram fs-5"></i></a>
        </div>
    </div>
    <hr class="my-4" style="border-color: #2a3b4d;">
    <div class="row">
        <div class="col-md-6">
            <p class="mb-0">© 2025 KnowledgeGraph Vis. Исследовательский
        проект.</p>
        </div>
        <div class="col-md-6 text-md-end">
            <a href="#" class="text-secondary">Политика конфиденциальности</a>
        </div>
    </div>
</footer>

<!-- Bootstrap &amp; Scripts --&gt;
&lt;script
src="https://cdn.jsdelivr.net/npm/bootstrap@5.2.3/dist/js/bootstrap.bundle.min.js"&gt;&lt;/sc
ript&gt;
&lt;script&gt;
    // Плавная прокрутка
    document.querySelectorAll('a[href^="#"]').forEach(anchor =&gt; {
        anchor.addEventListener('click', function (e) {
            e.preventDefault();
            const targetId = this.getAttribute('href');
            if(targetId === '#') return;
            const targetElement = document.querySelector(targetId);
            if(targetElement) {
                window.scrollTo({
                    top: targetElement.offsetTop - 80,
                    behavior: 'smooth'
                });
            }
        });
    });

    // Генерация графа для демо
    function createGraphNodes(containerId, nodeCount) {
        const container = document.getElementById(containerId);
        container.innerHTML = '';
        for(let i = 0; i &lt; nodeCount; i++) {
            const node = document.createElement('div');
            node.className = 'graph-node';
            node.style.width = Math.random() * 30 + 15 + 'px';
            node.style.height = node.style.width;
            node.style.left = Math.random() * 85 + '%';
            node.style.top = Math.random() * 85 + '%';
            container.appendChild(node);
        }
    }

    // Обработка формы
</pre>

```

```
document.getElementById('subscribe-form').addEventListener('submit',
function(e) {
    e.preventDefault();
    const emailInput = document.getElementById('email-input');
    if(emailInput.checkValidity()) {
        alert('Спасибо! Мы сообщим вам о запуске.');
        emailInput.value = '';
    } else {
        alert('Пожалуйста, введите корректный email адрес.');
    }
});

// Инициализация графов при загрузке
document.addEventListener('DOMContentLoaded', function() {
    createGraphNodes('hero-graph', 25);
    createGraphNodes('demo-graph', 35);
});

</script>
</body>
</html>
```