

УДК 004.413.2 + 930

EDN
[JGSFGV](#)

Проектирование агрегатора научно-технической информации по археологии с использованием методов искусственного интеллекта

П.Е. Артемьев*

Сибирский федеральный университет, пр. Свободный, 79, Красноярск, 660041, Россия

*E-mail: sanchello13pv@mail.ru

Аннотация. В данной статье рассматривается проблема систематизации и анализа археологических данных. Археология, как наука, сталкивается с вызовами, связанными с растущим объемом и разнообразием информации, что создает трудности в доступе и анализе данных. Актуальность проекта «Проектирование агрегатора научно-технической информации по археологии с использованием методов искусственного интеллекта» обусловлена необходимостью создания эффективных инструментов для автоматизации обработки и анализа информации, что позволит повысить точность и скорость исследований как в данной области, так и в смежных науках (гуманитарные, естественно-научные). Статья описывает цель работы, состоящую в разработке программного комплекса для сбора, структурирования и анализа данных из различных источников. Основное внимание уделяется использованию методов машинного обучения и обработки естественного языка для автоматизации процессов аннотирования и классификации данных. Данная работа представляет собой применение автоматизации в анализе и обработке научной информации в сфере археологии, предоставляя в конечном итоге исследователям необходимый инструмент для эффективной работы с данными, а также возможности широкому кругу пользователей в ознакомлении с этой предметной областью. Результаты и выводы исследования демонстрируют значимость интеграции новейших технологий и подходов в археологии, подчеркивая необходимость дальнейших исследований в данной области.

Ключевые слова: программная инженерия, археология, искусственный интеллект, анализ данных, клиент-серверное приложение.

Designing an aggregator of scientific and technical information on archeology using artificial intelligence methods

P.E. Artemev*

Siberian Federal University, 79 Svobodny pr., Krasnoyarsk, 660041, Russia

*E-mail: sanchello13pv@mail.ru

Abstract. This paper proposes the problem of systematization and analysis of archaeological data. Archaeology, as a science, is facing challenges related to the growing volume and variety of information, which creates difficulties in accessing and analyzing data. The relevance of the project "Designing an aggregator of scientific and technical information on archeology using artificial intelligence methods" is due to the need to create effective tools for automating information processing and analysis, which will improve the accuracy and speed of research both in this field and in related sciences (humanities, natural sciences). The article describes the purpose of the work, which is to develop a software package for collecting, structuring and analyzing data from various sources. The main focus is on the use of machine learning and natural language processing techniques to automate data annotation and classification processes. This work represents the application of automation in the analysis and processing of scientific information in the field of archaeology, providing researchers with the necessary tool for effective work with data, as well as opportunities for a wide range of users to familiarize themselves with this subject area. The results and conclusions of the study demonstrate the importance of integrating the latest technologies and approaches in archaeology, emphasizing the need for further research in this area.

Keywords: software engineering, archeology, artificial intelligence, data analysis, client-server application.

1. Введение

В последние десятилетия происходит значительный рост объема доступной информации, связанной с археологическими исследованиями, однако существующие базы данных и ресурсы часто являются фрагментированными и разрозненными. Это создает существенные трудности для специалистов в этой области, так как доступ к необходимой информации затруднен, а возможности для комплексного анализа ограничены.

Актуальность проекта «Проектирование агрегатора научно-технической информации по археологии с использованием методов искусственного интеллекта» обусловлена растущей объемностью и разнообразием данных в археологии. Научная новизна данной работы заключается в использовании современных технологий машинного обучения и обработки естественного языка для автоматизации процессов сбора и анализа археологической информации. Это позволит не только повысить эффективность работы исследователей, но и создать уникальный инструмент для анализа археологических источников, их классификации и сопоставления.

В качестве основы для анализа предметной области были найдены схожие проекты по своей реализации, концепции в предметной области [3, 7]; в других предметных областях [1, 4-6, 8].

Автором установлено, что в настоящее время в предметной области «археология» отсутствуют работы, посвящённые реализации агрегатора научной археологической информации с использованием новых методов машинного обучения и использования картографической информации, представленной в открытых данных по археологии, можно рассматривать данную работу как первую в данной области.

2. Постановка задачи (Цель исследования)

Целью работы является проектирование программного комплекса, который будет собирать, структурировать и обеспечивать доступ к научной информации из различных источников (отчёты, акты государственной историко-культурной экспертизы (ГИКЭ), публикации) по археологии.

Основными задачами разрабатываемой системы являются:

- Сбор данных. Автоматическое извлечение данных из открытых и специализированных научных источников;

- Обработка информации. Использование методов машинного обучения для классификации и аннотирования собранных данных;
- Анализ и поиск. Разработка алгоритмов для быстрого поиска и фильтрации информации на основе запросов пользователей;
- Реализация интерфейса. Создание удобного и информативного веб-интерфейса для пользователей.

3. Методы и материалы исследования

В таблице 1 представлен перечень всех используемых технологий в проекте. При подборе инструментов наибольшее внимание заслуживали те варианты, которые удобны в использовании и просты в освоении, а также имеют хорошую документацию. Необходимо учитывать взаимную совместимость и интеграцию инструментов, а также возможность с их помощью оптимизировать веб-приложение для повышения общей производительности. На основе этих принципов был выбран ряд технологий, которые наилучшим образом отвечают потребностям проекта.

Таблица 1. Стек используемых технологий.

№	Назначение	Инструмент
1	Язык программирования	JavaScript, HTML, CSS, JQuery, Python
2	Интегрированная среда разработки	WebStorm, PyCharm
3	Фреймворк кроссплатформенной разработки	Django
4	Система управления базами данных	PostgreSQL
5	Распределенная система управления версиями	Git
6	Систематизация задач и управление проектом	Trello
7	Диаграмма Ганта	Online Gantt

Для обработки и анализа данных использованы следующие методы: кластерный анализ, корреляционный анализ, регрессионный анализ, методы обработки естественного языка (NLP), машинное обучение; компьютерное зрение.

В результате реализации этих методов анализа данных будет получена качественно новая платформа для создания, анализа и экспертной оценки научной и научно-технической информации по археологии.

4. Полученные результаты

Клиентская часть представляет собой JavaScript-приложение с графическим интерфейсом, реализованным с использованием HTML, CSS, JS, Ajax. Клиентское приложение связывается с серверным приложением по протоколу HTTP [2].

Серверная часть состоит из веб-сервера и СУБД, которые взаимодействуют по протоколам TCP/UDP (LAN). СУБД осуществляет взаимодействие с базой данных при получении и обработке каких-либо запросов. веб-сервер включает в себя веб-контейнер, в который входят URL диспетчер, контроллер и компоненты модели данных. URL диспетчер обрабатывает обращения к тому или иному адресу, полученные от клиентской части. Контроллер обращается к СУБД и реализует ответ на полученные запросы. Компоненты модели данных являются представлением тех данных, которые были получены по запросу от базы данных (рисунок 1).

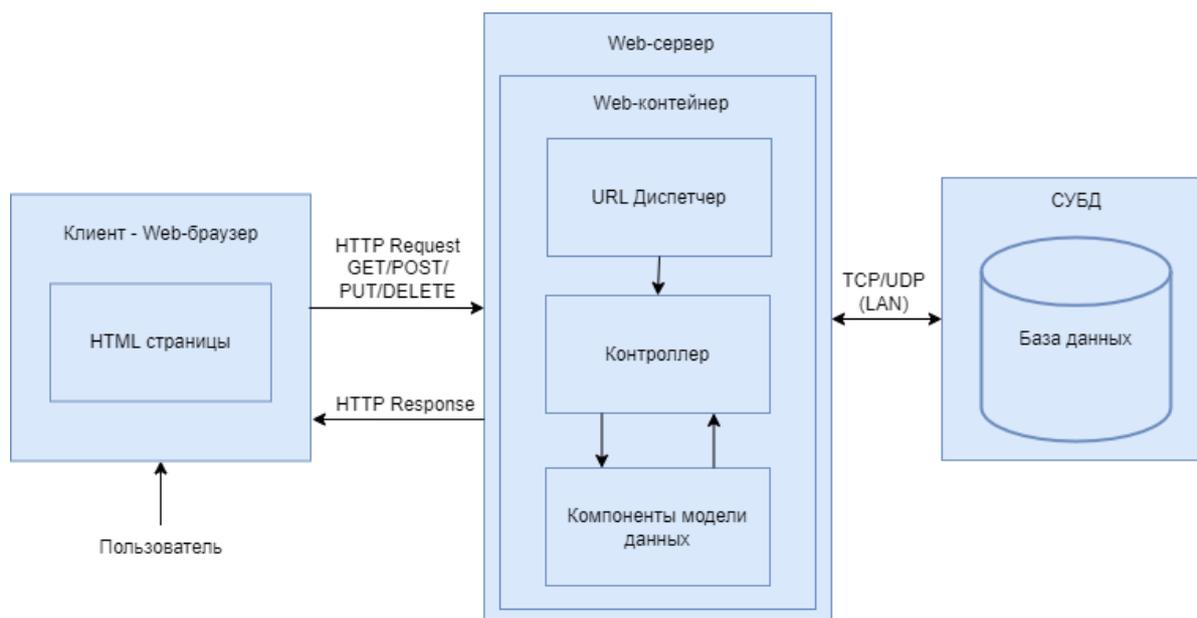


Рисунок 1. Архитектура приложения.

В состав системы должны входить следующие компоненты:

- Подсистема загрузки документов осуществляет загрузку файлов с локального устройства пользователя, либо производит скачивание документов с внешних источников.

- Подсистема обработки документов производит обработку всех загружаемых в систему документов – разбивает текст на структурные части документа, проводит оптическое распознавание текста на изображениях, извлекает и преобразует координаты и картографическую информацию для дальнейшего нанесения на интерактивную карту.
- Подсистема интеллектуального анализа текста проводит анализ агрегированной и формализованной информации, извлечённой из документов, и предоставляет его в удобном виде для отображения на пользовательском интерфейсе.

На основе требований к компонентному содержанию программной системы была составлена модульная схема приложения (рисунок 2).

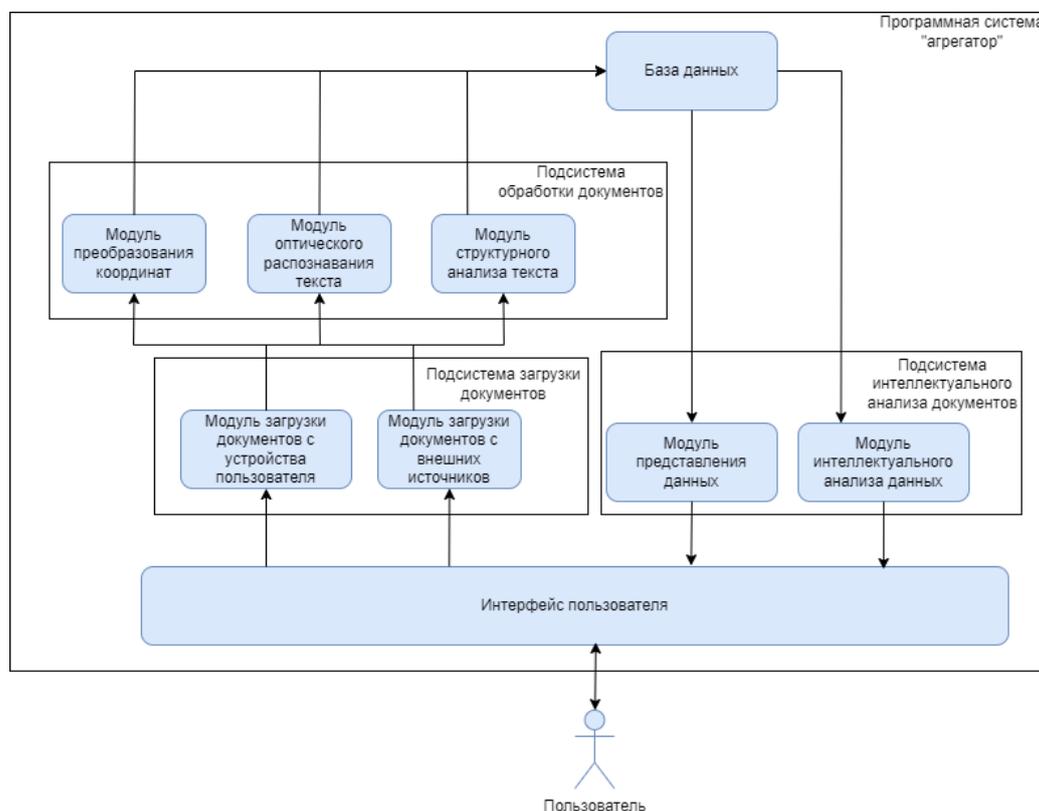


Рисунок 2. Модульная схема приложения.

5. Выводы

В результате работы проведен обзор существующих аналогов разрабатываемого приложения и их анализ, сформулированы требования к системе, определен технологический стек. После этого спроектированы база данных, серверная и клиентская

часть приложения. Спроектирован веб-сервис с учётом всех предъявляемых к нему требований.

В ходе реализации проекта выполнены следующие задачи:

- создан набор диаграмм, формализующий требования к системе;
- спроектирована база данных;
- спроектирована архитектура приложения в целом;
- спроектирована архитектура серверной части;
- спроектирована модульная архитектура приложения;
- спроектирован интерфейс разрабатываемого приложения в виде интерактивного макета;
- определен набор методов анализа данных и их назначение для работы с данными предметной области.

Данный сформированный набор документации призван значительно облегчить и организовать дальнейшую разработку программной системы.

Создание агрегатора открывает новые горизонты для дальнейшего развития в области археологического анализа и взаимодействия. Будущие исследования будут ориентированы на совершенствование алгоритмов машинного обучения и разработку дополнительных функций для глубокого анализа и интерпретации археологических данных.

Список литературы

1. Binaq, A. A. Design Platform Financial Aggregator in Agriculture / A. A. Binaq // Jurnal Sistem Cerdas. – 2019. – Vol. 2, no. 1. – P. 34-46. – DOI 10.37396/jsc.v2i1.16. – URL: <https://apic.id/jurnal/index.php/jsc/article/view/16/18> (дата обращения: 23.02.2025).
2. Богданенко, Д. А. Подходы к архитектурному проектированию веб-приложений / Д. А. Богданенко // Молодой ученый. – 2018. – № 9 (195). – С. 24-29. – URL: <https://moluch.ru/archive/195/48609/> (дата обращения: 23.02.2025).
3. Виноградская, М. Ю. Проектирование модели веб-сайта «археология» / М. Ю. Виноградская, Е. И. Шаронов // Дневник науки. – 2023. – № 1(73). – EDN PIARCA. – URL: https://elibrary.ru/download/elibrary_50437601_99179818.pdf (дата обращения: 23.02.2025).
4. Казакова, Е. Б. Единая информационная система и единый агрегатор торговли "Березка" в системе государственных закупок и их правовое регулирование / Е. Б.

- Казакова, А. Ю. Шишкова // Наука. Общество. Государство. – 2020. – Т. 8, № 2(30). – С. 162-173. – DOI 10.21685/2307-9525-2020-8-2-18. – EDN TRNFVY. – URL: https://elibrary.ru/download/elibrary_43054672_51584298.pdf (дата обращения: 23.02.2025).
5. Колосок, И. Н. Demand Response агрегатор как бизнес-процесс в условиях цифровизации энергетики / И. Н. Колосок, Е. С. Коркина // Релейная защита и автоматизация. – 2021. – № 4(45). – С. 62-67. – EDN FSLLSW. – URL: https://elibrary.ru/download/elibrary_47270705_17232797.pdf (дата обращения: 23.02.2025).
6. Лапина, Т. И. Разработка и реализация интернет-платформы агрегатора автосервисов / Т. И. Лапина, В. А. Сарычев, В. Ю. Цевка // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. – 2022. – Т. 12, № 2. – С. 40-58. – DOI 10.21869/2223-1536-2022-12-2-40-58. – EDN PISNZI. – URL: https://elibrary.ru/download/elibrary_49350797_91411111.pdf (дата обращения: 23.02.2025).
7. Сомов, А. К. Цифровая платформа Енисейской Сибири “Сибириана” / А. К. Сомов, Е. Р. Брюханова, О. А. Антамошкин, Т. С. Плешкова // Журнал Сибирского федерального университета. Серия: Гуманитарные науки. – 2024. – Т. 17, № 9. – С. 1782-1789. – EDN RMZJPT. – URL: https://elibrary.ru/download/elibrary_74508942_37595956.pdf (дата обращения: 23.02.2025).
8. Smirnov, D. Pharmacy Aggregator Web Application / D. Smirnov. – Bachelor of Engineering Thesis. – 2019. – 57 p. – URL: <https://www.theseus.fi/bitstream/handle/10024/266434/DaniilSmirnov%20Thesis%20Final.pdf?sequence=2&isAllowed=y> (дата обращения: 23.02.2025).