# Техническое задание

# **Общие сведения**

**Название системы.** Система интеллектуальной обработки и классификации веб-сайтов с последующей интеграцией результатов в образовательную среду.

**Основания для создания системы.** Необходимость автоматизации процессов сбора, предобработки, классификации и поиска по научным статьям и другим веб-ресурсам для упрощения доступа студентов и преподавателей к релевантной информации. Система направлена на повышение эффективности образовательного процесса за счет минимизации временных затрат на поиск и анализ данных.

# **Цели и назначение создания автоматизированной системы**

Повышение качества и скорости поиска научных материалов для студентов и преподавателей путем внедрения интеллектуальной системы анализа и классификации веб-сайтов.

1. Автоматизация парсинга веб-страниц для извлечения заголовков, описаний и содержимого.
2. Классификация текстового контента по 40 категориям из сферы компьютерных наук с помощью нейронных сетей.
3. Формирование краткой выжимки из содержимого веб-страниц для улучшения пользовательского опыта.
4. Реализация семантического поиска по базе данных для быстрого нахождения релевантных источников.
5. Интеграция системы в Telegram-бот для удобства использования.

# **Характеристика объектов автоматизации**

Объекты автоматизации:

1. Процесс сбора данных с веб-страниц (парсинг).
2. Предобработка текстовых данных (удаление html-тегов, пунктуации и спецсимволов, удаление стоп-слов и лемматизация).
3. Классификация текстового контента с использованием нейронной сети.
4. Формирование краткой выжимки из текстового контента.
5. Взаимодействие пользователей с системой через Telegram-бот.

# **Требования к автоматизированной системе**

Функциональные характеристики:

**Парсер**: Должен извлекать заголовки, описания и основное содержимое веб-страниц. Поддержка работы с динамическим контентом.

**Предобработка данных**: Удаление пунктуации, стоп-слов и html-тегов, нормализация, лемматизация, перевод текста на русский язык.

**Модель нейронной сети:** Многоклассовая классификация текстов по 40 категориям. Возможность переобучения модели на новых данных.

**База данных**: Хранение URL, заголовков, описаний, меток классификации и краткой выжимки.

**Telegram-бот**: Интерфейс для отправки запросов и получения релевантных источников с краткими выжимками.

**Требования к надежности:**

Система должна обеспечивать стабильную работу при высоких нагрузках.

Логирование ошибок и мониторинг состояния системы.

Резервное копирование данных каждые 2 недели.

Требования к эргономике и техническому обслуживанию:

Простота использования Telegram-бота для конечных пользователей.

Наличие подробной документации и руководства пользователя.

**Требования к совместимости**:

Совместимость с современными операционными системами (Windows, Linux, macOS).

**Требования к производительности:**

Максимальная задержка ответа на запрос пользователя: не более 5 секунд.

Обработка не менее 30 запросов в минуту.

**Используемые технологии и инструменты:**

Парсинг: Selenium, BeautifulSoup.

Предобработка текста: NLTK, spaCy.

Модель нейронной сети: TensorFlow/Keras (BERT или LSTM).

Telegram-бот: Aiogram.

База данных: PostgreSQL или MongoDB.

# **Состав и содержание работ по созданию автоматизированной системы**

**Анализ требований и проектирование системы:**

* Продолжительность: 2 месяца.
* Результаты: Функциональные и технические требования, архитектурное решение.

**Реализация системы:**

* Продолжительность: 3 месяца.
* Результаты: Работоспособные компоненты системы (парсер, модуль предобработки, модель нейронной сети, база данных, Telegram-бот).

**Тестирование и внедрение:**

* Продолжительность: 1 месяц.
* Результаты: Протестированная и готовая к эксплуатации система.

**Эксплуатация и сопровождение:**

* Результаты: Обеспечение бесперебойной работы системы, регулярное обновление и улучшение.

# **Порядок разработки автоматизированной системы**

Порядок разработки системы включает:

1. Анализ требований и проектирование системы.
2. Разработка компонентов системы (парсер, предобработка, модель, база данных, бот).
3. Тестирование системы на реальных данных.
4. Внедрение системы и подготовка к эксплуатации.
5. Обеспечение технической поддержки и регулярное обновление.

# **Порядок контроля и приемки автоматизированной системы**

Контроль и приемка осуществляются руководителем проекта на каждом этапе. Приемочные испытания проводятся после завершения реализации.

Критерии приемки:

1. Соответствие функциональным и техническим требованиям.
2. Надежность и безопасность системы.
3. Производительность системы.
4. Качество и полнота документации.

# **Требования к составу и содержанию работ по подготовке объекта автоматизации к вводу автоматизированной системы в действие**

* Подготовка сервера для размещения системы.
* Установка необходимых программных модулей.
* Настройка базы данных и сервера.
* Проведение тестирования системы перед запуском.

# **Требования к документированию.**

* Полный комплект эксплуатационной и программной документации.
* Описание всех функций и характеристик системы.
* Руководство пользователя, административное руководство, техническое описание системы.

# **Источники разработки**

Источником финансирования проекта является разработчик системы. Технические и методические материалы предоставляются из открытых источников информации на основе анализа существующих аналогов и лучших практик в области информационных систем для образования.