

Описание функциональных характеристик программного обеспечения

Наименование ПО

Руконтекст 2.0 — система поиска заимствований и проверки текстовых документов на оригинальность (антиплагиат-система).

Класс ПО

Прикладное программное обеспечение — средства управления процессами / системы проверки оригинальности текстов.

Назначение

Руконтекст 2.0 предназначен для автоматизированной проверки текстовых документов на наличие заимствований (плагиата) с использованием алгоритмов машинного обучения и полнотекстового поиска.

Целевые потребители: образовательные учреждения, научные организации, издательства, коммерческие компании, государственные структуры.

Функциональные характеристики

1. Проверка документов на заимствования

- Загрузка документов в форматах: DOC, DOCX, ODT, PDF (включая отсканированные), RTF, TXT
- Автоматическое распознавание текста (OCR) для отсканированных документов с поддержкой русского и английского языков
- Конвертация документов во внутренний формат с сохранением структуры
- Анализ текста с использованием алгоритмов: нечёткое сравнение (MoreLikeThis), машинное обучение (Accord.MachineLearning), морфологический анализ (лемматизация для русского и английского языков)
- Поиск заимствований по собственным индексным коллекциям организации
- Проверка по внешним открытым источникам (при наличии подключения)
- Кросс-языковая проверка (русский, английский) — опционально

2. Детектор генеративного текста (AI-текста)

- Обнаружение текстов, сгенерированных нейронными сетями (GPT и др.)
- Реализовано на базе PyTorch и Hugging Face Transformers
- Поддержка работы на CPU и GPU (NVIDIA CUDA)
- Интеграция в процесс проверки документа

3. Формирование отчётов

- Генерация PDF-отчётов о результатах проверки с детализацией найденных заимствований
- Отображение процента оригинальности, цитирования и заимствований
- Визуальная разметка заимствованных фрагментов с указанием источников
- Экспорт результатов в формат Excel (XLSX)

4. Управление индексными коллекциями

- Добавление документов в индекс для последующего поиска заимствований
- Поддержка нескольких индексных коллекций (по подразделениям, тематикам)
- Резервное копирование и восстановление индексов (снапшоты)
- Импорт индексных коллекций с центрального сервера

5. Управление пользователями и доступом

- Ролевая модель доступа: поддержка (Support), менеджер (Manager), преподаватель (Teacher), студент (Student), модератор (Moderator), супервизор (Supervisor), эксперт (Expert)
- Регистрация и аутентификация пользователей
- Интеграция с системами управления обучением по протоколу LTI 1.1 и LTI 1.3 JWT-аутентификация для API

6. Система подписок и лимитов

- Управление подписками организаций с ограничениями по количеству проверок
- Дневные и периодические лимиты проверок
- Тикетная система — возможность выдачи одноразовых билетов на проверку
- Контроль расхода проверок с уведомлением при приближении к лимиту

7. Организационная структура

- Поддержка иерархической структуры подразделений организации
- Привязка пользователей к подразделениям
- Агрегация статистики по организационным единицам
- Справочник должностей и ролей

8. Машинный перевод (опционально)

- Автоматический перевод текстов между языками
- Реализовано на базе PyTorch и Hugging Face Transformers
- Требуется GPU (NVIDIA CUDA)
- Используется для кросс-языковой проверки

9. Выравнивание параграфов (опционально)

- Параллельное выравнивание текстов на разных языках
- Применяется при кросс-языковом поиске заимствований
- Реализовано на базе PyTorch
- Требуется GPU (NVIDIA CUDA)

10. Уведомления

- Отправка email-уведомлений о завершении проверки
- Очередь отправки через брокер сообщений (RabbitMQ)
- Поддержка SMTP-протокола (через MailKit)
- Уведомления в реальном времени через WebSocket (SignalR)

11. API для интеграции

- REST API для программного взаимодействия с системой
- JWT-авторизация запросов
- Операции: загрузка документов, запуск проверки, получение результатов, управление пользователями
- Поддержка интеграции по протоколу LTI 1.3 с LMS-системами

12. Многоязычный интерфейс (индивидуально для каждого клиента)

- Русский язык (основной)
- Английский язык
- Белорусский язык
- Узбекский язык
- Динамическая локализация из базы данных

13. Администрирование

- Веб-панель администратора
- Мониторинг состояния сервисов
- Управление конфигурацией через etcd
- Профилирование производительности (MiniProfiler)
- Структурированное журналирование (NLog)

Системные требования

Минимальные требования (сервер)

Параметр	Значение
Процессор	x86-64 (AMD64), 16 ядер
Оперативная память	64 ГБ
Дисковое пространство	512 ГБ SSD
Операционная система	Linux (Ubuntu 22.04+, Astra Linux, и др.)
Docker	24.0+ с Docker Compose v2

Рекомендуемые требования

Параметр	Значение
Процессор	x86-64, 32+ ядер
Оперативная память	256+ ГБ

Параметр	Значение
Дисковое пространство	1+ ТБ SSD
Операционная система	Linux (Ubuntu 22.04+, Astra Linux, и др.)
Docker	24.0+ с Docker Compose v2
GPU (опционально)	NVIDIA с поддержкой CUDA 11.7+

Требования к клиенту (браузер)

- Любой современный веб-браузер: Яндекс.Браузер, Chromium, Firefox, Safari
- Разрешение экрана: 1280x720 и выше

Совместимость с ОС из реестра российского ПО

Система поставляется в виде Docker-контейнеров и совместима с любой Linux-операционной системой, поддерживающей Docker 24.0+:

- Astra Linux Special Edition / Common Edition
- ALT Linux (Альт)
- Ubuntu (в т.ч. для тестирования)

Ограничения

- Для работы требуется серверное оборудование (не предназначено для рабочих станций)
- GPU-зависимые компоненты (переводчик, выравнивание параграфов) требуют NVIDIA GPU
- Минимальное количество документов в индексе для эффективной работы поиска: 3