

Промежуточный отчет по программному проекту

1. Основные планы и задачи разработчика Пуховой А.И.

1.1 Краткое описание проекта:

«Веб-приложение для комплексной автоматизации процессов поискового продвижения веб-сайтов» («SEO Мастер») — это высокоавтоматизированная, микросервисная платформа, предназначенная для комплексного управления и оптимизации SEO-продвижения веб-сайтов. Платформа обеспечивает независимость от дорогих внешних SEO API, внедряет механизм Human-in-the-Loop (HITL) для контроля критических изменений и фокусируется на бизнес-ориентированной аналитике (FF-Score, E-E-A-T Score).

Цель проекта: разработка веб-приложения для автоматизированного SEO-аудита и долгосрочного управления продвижением сайтов с использованием событийной архитектуры и аналитических метрик.

Название проекта:

Веб-приложение для комплексной автоматизации процессов поискового продвижения веб-сайтов

Цель работы (Пухова А.И.):

Разработка ключевых микросервисов, отвечающих за сбор и анализ данных (Audit Service), семантический анализ и генерацию контента (Semantic & Content Service), а также сбор внешних данных и отчетность (Reporting Service). Обеспечение надежной работы аналитического ядра платформы.

Краткое описание задач:

1. Разработка и реализация Audit Service (Краулер, Индексатор Ссылок, Анализаторы CWV/Schema).
2. Разработка и реализация Semantic & Content Service (LLM Адаптер, Расчет FF-Score, E-E-A-T Анализ, Контент-Анализ).
3. Разработка и реализация Reporting Service (Сборщики GSC/GA4/Yandex, Расчет Метрик, Генерация Отчетов PDF/CSV).
4. Обеспечение асинхронного взаимодействия между микросервисами через событийную модель (Celery/RabbitMQ).

1.2 Планы и этапы выполнения проекта (Специфично для Разработчика 2)

| Микросервис | Ключевые модули | Технологии | Статус/Сроки |
|--------------|---|-------------------|-------------------------|
| Общие задачи | Формализация логики FF-Score и HITL-механизма, включая реализацию событийной модели с использованием событий CrawlCompleted и FFScoreRecalculated | Celery / RabbitMQ | 17.12.2025 – 10.01.2026 |

| Микросервис | Ключевые модули | Технологии | Статус/Сроки |
|-------------------|--|---|-------------------------|
| Audit Service | Краулер (<code>technical_audit.py</code>), Индексатор Ссылок (<code>backlink_indexer.py</code>), Анализаторы CWV/Schema | Python, Scrapy, Playwright, PostgreSQL | 11.01.2026 – 25.02.2026 |
| Semantic Service | LLM Адаптер (<code>llm_client.py</code>), FF-Score (<code>ff_score_calculator.py</code>), Е-Е-А-Т Анализ (<code>eeat_analyzer.py</code>), Контент-Анализ | Python, LLM API (OpenAI/Gemini), NLTK/spaCy | 26.02.2026 – 20.03.2026 |
| Reporting Service | Сборщики GSC/GA4/Yandex, Расчет Метрик (<code>calculator.py</code>), Генерация Отчетов PDF/CSV | Python, pandas, pdfkit, GSC/GA4 API | 01.03.26 – 25.03.26 |

2. Используемый технологический стек и его обоснование

2.1 Перечень используемых технологий

| Технология/Инструмент | Назначение | Обоснование |
|-------------------------|--|---|
| Python (FastAPI/Django) | Основной язык для Backend-логики микросервисов | Высокая скорость разработки, богатая экосистема для NLP/SEO (Scrapy, pandas) |
| PostgreSQL (с JSONB) | Основное хранилище данных и собственного индекса ссылок | Надежность, поддержка сложных запросов, JSONB для гибкого хранения связанных данных |
| Celery / RabbitMQ | Очередь задач для асинхронного выполнения долгих процессов | Необходимость для краулинга и LLM-генерации, которые являются долгими операциями |
| Scrapy / Playwright | Краулинг и рендеринг JS/SPA контента | Scrapy для эффективности, Playwright для обработки современного JS-рендеринга |
| LLM API (OpenAI/Gemini) | Генерация контента и метаданных, семантический анализ | Обеспечение функционала LLM-генерации черновиков и Е-Е-А-Т анализа |
| pandas / pdfkit | Сбор, обработка данных и генерация отчетов (Reporting Service) | pandas для эффективной работы с данными, pdfkit для генерации PDF-отчетов |
| Git, GitHub | Контроль версий исходного кода проекта и хранение репозитория | Позволяют отслеживать изменения в коде, управлять версиями проекта и обеспеч |

| Технология/Инструмент | Назначение | Обоснование |
|-----------------------|--|--|
| | | чивать сохранность результатов разработки |
| Visual Studio Code | Интегрированная среда разработки для реализации backend-компонентов проекта, включая разработку краулеров, анализаторов, модулей семантического и контентного анализа, а также средств формирования отчёtnости | Бесплатная кроссплатформенная среда разработки с широкой экосистемой расширений для Python, работы с краулерами и NLP-модулями, контейнеризации (Docker), системы контроля версий Git, а также сопутствующих инструментов аналитики и отчёtnости |

2.2 Обоснование выбранного технологического стека

Выбор **Python** как основного языка обусловлен его доминированием в области **Data Science**, **NLP** и **Web Scraping**. Библиотеки **Scrapy**, **pandas**, **NLTK/spaCy** являются критически важными для реализации **Audit** и **Semantic** сервисов.

Использование **PostgreSQL** с типом данных **JSONB** позволяет эффективно хранить как структурированные данные (например, метаданные страниц), так и неструктурированные сырье данные краулинга, что соответствует требованиям к **Хранилищу Ссылок** в **Audit Service**.

Celery и **RabbitMQ** выбраны для реализации **Event-Driven Architecture (EDA)**. Это необходимо для асинхронного выполнения ресурсоемких задач, таких как полный крауллинг срендерингом JS и сложные расчеты **FF-Score**, предотвращая блокировку основного API.

Playwright обеспечивает возможность полного рендеринга страниц с **JavaScript**, что является ключевым требованием для **Audit Service** при анализе современных **SPA** (Single Page Applications) и корректной оценке **Mobile-First Indexing**.

Использование **Git** и **GitHub** обусловлено необходимостью централизованного хранения исходного кода и контроля версий проекта. Это позволяет фиксировать этапы разработки, отслеживать изменения в кодовой базе и обеспечивать согласованную работу с архитектурными и программными артефактами проекта.

Visual Studio Code используется в качестве основной среды разработки благодаря встроенной поддержке языка **Python**, удобным инструментам отладки и интеграции с системой контроля версий. Применение данных **IDE** повышает эффективность разработки и упрощает сопровождение кода на всех этапах реализации проекта.

3. Критерии оценивания работы (Пухова А.И.)

| Критерий | Описание |
|------------------------------------|---|
| Функциональность Audit Service | Корректная работа JS-краулера, полнота индексации ссылок, валидация CWV и Schema.org. |
| Функциональность Semantic Service | Точность расчета FF-Score и E-E-A-T Score, корректная работа LLM-адаптера и генерации черновиков. |
| Функциональность Reporting Service | Успешное подключение к GSC/GA4/Yandex API, корректный расчет Cost-Efficiency, генерация отчетов PDF/CSV. |
| Соблюдение архитектуры | Строгое следование принципам DDD и EDA, корректная публикация и потребление доменных событий (CrawlCompleted, FFScoreRecalculated). |
| Качество кода | Соответствие Clean Architecture, покрытие тестами, читаемость и поддерживаемость кода. |

4. Особые пометки

Настоящий документ описывает только индивидуальную часть проекта, разрабатываемую Пуховой А.И.

В рамках командного проекта распределение ответственности следующее:

Зобов А.А. (Разработчик 1) отвечает за:

- Разработку и реализацию Management Service
- Разработку и реализацию API Gateway / Frontend Service
- Разработку и реализацию Client API Gateway и адаптеров для клиентов
- Инициализацию и настройку общей инфраструктуры проекта (Docker, Celery/RabbitMQ, миграции БД)

Пухова А.И. (Разработчик 2) отвечает за:

- Разработку Audit Service (краулер, индексатор, анализ Core Web Vitals)
- Разработку Semantic Service (LLM-анализ, расчет E-E-A-T Score и FF-Score)
- Реализацию Reporting Service (формирование отчетов, интеграция с внешними API)
- Участие в разработке Client API Gateway (адаптеры)

Интеграция компонентов и совместное тестирование системы будет выполняться обоими разработчиками в период с 16.04.26 по 20.04.26 согласно графику проекта.