

Техническое задание

Проект: Система анализа отзывов с классификацией тональности (по ссылке на товар)

Источники: DNS / M.Видео / Ozon / Wildberries / Ситилинк и др. (минимум 3)

1. Цель работы

Разработать веб-приложение, которое по ссылке на товар в интернет-магазине собирает отзывы, анализирует их с помощью ИИ и отображает результаты на сайте: долю положительных/нейтральных/негативных отзывов, основные плюсы и минусы, примеры отзывов.

2. Поддерживаемые источники

- Минимум 3 источника (магазина) из списка: DNS, M.Видео, Ozon, Wildberries, Ситилинк (или другие по согласованию).
- Пользователь вставляет URL товара, система автоматически определяет магазин по домену.

3. Основной сценарий пользователя

- Пользователь вставляет ссылку на товар и нажимает кнопку «Проанализировать».
- Система собирает отзывы по товару (пагинация при наличии).
- Отзывы сохраняются в базу данных и передаются в модуль анализа тональности.
- Пользователь видит страницу результата: общий итог, графики/цифры, топ-темы, примеры отзывов.

4. Сбор отзывов (получение данных)

Сбор отзывов реализуется на серверной стороне. Допускается один из вариантов:

- Официальный/публичный API магазина (если доступен и легален для использования).
- Парсинг HTML страниц отзывов (requests + HTML-парсер) — для учебного проекта.

Требование безопасности и корректности: не выполнять агрессивный скрейпинг. Обязательны ограничение частоты запросов (rate limit), таймауты, повторные попытки и логирование ошибок. Для демонстрации достаточно собрать, например, до 200-500 отзывов на товар.

5. ИИ-анализ (тональность и сводка)

- Использовать API ИИ (например: OpenAI / YandexGPT / аналог) через backend.
- Для каждого отзыва определить метку: позитивный / нейтральный / негативный.

- Дополнительно сформировать сводку по товару: основные плюсы/минусы (списками), частые проблемы.
- Сохранять результаты анализа в БД, чтобы не пересчитывать при повторном открытии.

6. Отображение результата на сайте

- Карточка товара: название (если доступно), магазин-источник, ссылка, дата анализа.
- Сводные показатели: количество отзывов, доля тональностей (в %), средняя оценка (если доступна).
- Список «Плюсы» и «Минусы» (сформированы ИИ).
- Примеры отзывов по категориям (3–5 позитивных и 3–5 негативных).
- Фильтрация отзывов по тональности.

7. Административная часть (минимум)

- Авторизация администратора.
- Просмотр истории запусков анализа (какие URL анализировались, сколько отзывов, статус).
- Просмотр логов сборщика (ошибки источников).
- Возможность повторно запустить сбор/анализ для выбранного товара.

8. Хранимые данные (MySQL, минимум)

- sources: id, name, domain.
- products: id, source_id, url, title (опц.), created_at, updated_at.
- reviews: id, product_id, author (опц.), rating (опц.), text, date (опц.), raw_id/url (опц.).
- review_analysis: review_id, sentiment (pos/neu/neg), score/confidence (опц.), tags (опц.).
- product_summary: product_id, pros_text, cons_text, overall_summary, analyzed_at.
- jobs/logs: job_id, status, message, started_at, finished_at.

9. Технические требования

- Frontend: HTML/CSS/JS или React (страница ввода ссылки + страница результата).
- Backend: Node.js или PHP (рекомендуется выделить модуль «сборщик» и модуль «анализатор»).
- База данных: MySQL.
- Обязательная обработка ошибок: недоступная страница, капча/блокировка, пустые отзывы, таймауты, ошибки ИИ API.
- README с инструкцией запуска и примерами ссылок для тестирования.

10. Ограничения

- Оплата/заказы/корзина не реализуются.
- Достаточно ограниченного количества источников (≥ 3) и ограниченного объема отзывов на один товар.
- Если источник не позволяет получить отзывы (капча/блок), система должна корректно показать сообщение об ошибке и сохранить лог.

11. Результат работы

- Рабочее веб-приложение: вставка URL → сбор отзывов → анализ → страница результата.
- Демонстрация минимум на 3 магазинах (3 разных домена).
- Исходный код + база данных с сохраненными результатами анализа.

12. Критерии оценки

- Корректное определение магазина по ссылке и успешный сбор отзывов.
- Наличие классификации тональности (pos/neu/neg) и сохранение результатов.
- Наглядная страница результата (проценты, плюсы/минусы, примеры).
- Стабильность (обработка ошибок, rate limit, логирование).
- Качество интерфейса и адаптивность.