

KLT 5.01 ALPHA

Техническое задание и архитектура программы

Переход KLT 5.0 beta к KLT 5.01 Alpha: данные, модули, API, отчеты, тесты

Статус: техническая спецификация исходной alpha-сборки. **Цель** - единый контур обработки больших пакетов документов: сметы, договоры, тендеры, бухгалтерия, банк, логистика, открытые закупочные данные.

1. Назначение и пользователи

Пользователь	Сценарий
Сметчик / инженер	Проверка сметных строк, объемов, материалов, наложений и неполных оснований.
Финансовый аналитик	Проверка источников финансирования, кассовых разрывов, платежей, кредитной нагрузки.
Бухгалтер / аудитор	Сверка договоров, актов, платежей, проводок, запасов, дебиторки и кредиторки.
Заказчик / ведомство	Предварительный риск-скрининг закупки, проекта или исполнения контракта.
Антикоррупционный аналитик	Выявление красных флагов без формулирования обвинения.

KLT 5.01 Alpha: контур проектно-финансового лямбда-аудита



Назначение: найти точки, где проект не замыкается в доказанную, реализуемую и экономически обоснованную структуру.

Рисунок 1. Архитектура обработки данных KLT 5.01 Alpha

2. Поддерживаемые входные форматы

Формат	Статус в KLT 5.01 Alpha	Назначение
ZIP / RAR	Alpha: ZIP обязательно; RAR через внешний unrar/bsdtar	Массовая загрузка пакетов документов.
XLSX / CSV	Alpha: обязательно	Сметы, реестры платежей, финансовые таблицы, логистика.
DOCX / TXT / MD	Alpha: обязательно	Договоры, ТЗ, пояснительные записки, учетная политика.
PDF	Alpha: обязательно для текстового PDF; OCR как расширение	Тендерная документация, сметы, акты, банковские документы.
XML / JSON	Alpha: обязательно для машинных выгрузок	ЕИС/закупочные данные, внутренние базы, API.
DBF	Alpha: адаптер для банковской/учетной отчетности	Форма 101 и иные табличные выгрузки, где применимо.
DXF / IFC / BIM	Roadmap	Связь чертежных объектов и сметных позиций.
DWG	Roadmap через DWG->DXF	Закрытый формат, требуется внешний конвертер или SDK.

3. Модульная структура проекта

Папка	Модуль	Назначение
klt501/core	lambda_engine.py	Расчет lambda_i, Delta_i, Auth_i, глобального функционала и весов.
klt501/core	pix_peaks.py	Единая модель PIX/PEAKS и статусов согласованности.
klt501/core	project_graph.py	Граф доказательной связности: документы, позиции, ребра, атрибуты.
klt501/adapters	procurement_adapter.py	ЕИС, извещения, НМЦК, протоколы, контракты, исполнение.
klt501/adapters	estimate_adapter.py	Локальные сметы, сводные расчеты, позиции, ресурсы, объемы.
klt501/adapters	accounting_adapter.py	Баланс, ОФР, проводки, запасы, дебиторка, кредиторка.
klt501/adapters	bank_adapter.py	Платежи, выписки, гарантии, кредиты, графики.
klt501/adapters	logistics_adapter.py	Поставки, склады, маршруты, ТТН, списания.
klt501/audit	gap_detector.py	Разрывы между документами и позициями.
klt501/audit	overlap_detector.py	Дубли и наложения работ, материалов, платежей.
klt501/audit	anticorruption_flags.py	Красные флаги без юридического обвинения.
klt501/reports	lambda_audit_docx.py	Фирменный бланк итогового lambda-чек-аудита.
klt501/reports	risk_register.py	Реестр проблемных позиций и рекомендаций.

4. Нормализованная модель данных

Сущность	Ключевые поля
Document	id, type, source_file, hash, pages, date, organization, sign_status
PIX	id, document_id, locator, text, amount, quantity, unit, account, contract_ref, date
PEAKS	pix_id, status, evidence_score, consistency_score, risk_status, reviewer_note
Edge	from_pix, to_pix, relation_type, confidence, evidence_link
Issue	id, pix_id, category, severity, amount, lambda_defect, explanation
Recommendation	issue_id, action_type, request_documents, recalc, exclude, confirm, legal_review
AuditReport	project_id, metrics, risk_register, diagrams, signatures, export_files

5. Pipeline обработки

Шаг	Операция	Контроль качества
1	Загрузка архива или файлов	Вычисление hash, журнал загрузки, проверка формата.
2	Распаковка и классификация документов	Тип документа, источник, дата, версия, организация.
3	Извлечение таблиц и текста	Проверка количества строк, валюты, единиц измерения, OCR-флаг.
4	Формирование PIX	Каждая позиция получает источник и локатор: файл/страница/строка.
5	Назначение PEAKS	Статус подтвержденности, противоречия, дублирования, слабого основания.
6	Построение графа	Проверка связей: смета -> договор -> акт -> платеж -> учет -> факт.
7	lambda-расчет	Локальные и глобальные метрики, веса, пороги.
8	Детекторы риска	Gap, Overlap, Redundancy, WeakEvidence, Overprice, Underprice, Collapse, ACR.
9	Формирование отчета	DOCX/PDF, реестр позиций, диаграммы, подпись проверяющего.

6. CLI, API и интерфейс

CLI пример:

```
python -m klt501.app --input project_package.zip --profile construction_procurement --out audit_output --lang ru --checker "ФИО / логин"
```

Endpoint	Метод	Назначение
/api/v1/audits	POST	Создать проверку и загрузить пакет.
/api/v1/audits/{id}/status	GET	Статус обработки, ошибки, прогресс.
/api/v1/audits/{id}/issues	GET	Реестр проблемных позиций.
/api/v1/audits/{id}/report.docx	GET	Скачать итоговый DOCX.
/api/v1/audits/{id}/report.pdf	GET	Скачать итоговый PDF.
/api/v1/profiles	GET/POST	Профили весов и нормативных требований.

7. Требования к трехязычному интерфейсу

Язык	Что локализуется
RU	Основной интерфейс, документы для российских ведомственных пилотов, реестр рисков.
EN	Инвесторские и международные демонстрации, API documentation, export labels.
ZH	Интерфейс и отчеты для китайских текстовых и проектных кейсов; словарь терминов выносится в отдельный ресурс.

8. Отчетность и фирменный бланк

- Каждый отчет должен содержать ID проверки, дату, перечень файлов, hash-паспорта, проверяющего и электронный адрес.
- Реестр проблемных позиций должен указывать точное место: документ, страница, таблица, строка, позиция, сумма.
- Антикоррупционный раздел формулируется как риск-флаги и рекомендации к проверке, без обвинения в совершении правонарушения.
- Для каждой рекомендации указывается действие: запросить документ, пересчитать, исключить, подтвердить, вынести на юридическую проверку.

9. Тесты и критерии приемки Alpha

Тест	Критерий приемки
selftest core	Расчет lambda_i, Delta_i, Auth_i проходит на тестовом наборе без ошибок.
batch ZIP	Архив с 20+ файлами распакован, классифицирован, обработан.
table extraction	Не менее 95% строк демо-сметы попали в PIX с локаторами.
graph integrity	Все основные цепочки связей строятся или получают явный статус missing.
risk register	Для каждой найденной проблемы есть категория, позиция, сумма, объяснение, рекомендация.
report render	DOCX/PDF отчет читается, таблицы не ломаются, черных квадратов нет.
language QA	RU/EN/ZH строки интерфейса присутствуют и не смешиваются.

10. Риски реализации и roadmap

Риск	Решение
Сканированные PDF без слоя текста	OCR-модуль и явный флаг доверия к извлечению.
Разнородные сметные формы	Профили извлечения и ручная верификация на пилоте.
DWG как закрытый формат	Маршрут DWG->DXF или внешний SDK.
Неполные открытые данные	Режим неполного доказательного графа с пометкой MissingEvidence.
Юридическая чувствительность	Формулировки только как аналитические признаки и рекомендации к проверке.