

Руководство пользователя

[English](#) · [Описание](#) · [ОКТМО](#) · [Справочник](#)

Содержание

- [1. Запуск](#)
- [2. Рабочие папки](#)
- [3. Как устроен экран команды](#)
- [4. Команды](#)
- [5. Обслуживание](#)
- [6. Порядок работы](#)

Программа работает окном: команды выбираются мышью, столбцы задаются буквами Excel, результат уходит в отдельную папку. Править файлы настроек для обычной работы не требуется.

1. Запуск

`launcher_gui.cmd` в корне проекта — обычный способ. `Start.exe` делает то же самое.

Launcher находит portable-окружение (`runtime\pythonw.exe`), затем `runtime\python.exe`, затем `runtime\python\`, затем системный `py -3.12`) и поднимает локальный интерфейс на `127.0.0.1:8080` в отдельном окне.

Если окно не появилось, задайте `AUDION_GUI_CONSOLE=1` и запустите launcher снова — он откроется в консоли с сообщениями об ошибках. Подробности — в [Установке](#).

2. Рабочие папки

Проект работает с шестью управляемыми папками:

Папка	Что в ней
<code>input</code>	Исходные файлы
<code>output</code>	Результаты
<code>report</code>	JSON-отчёты прогонов
<code>logs</code>	Журналы операций
<code>workspace</code>	Промежуточные файлы
<code>release</code>	Архивы результатов

Источник и **Назначение** в верхней части окна переопределяют `input` и `output` для текущей работы. Путь можно закрепить булавкой — тогда он всегда будет первым в списке недавних. История путей хранится в `config/path_history.json`.

Кнопки работы с исходниками:

Кнопка	Действие
Добавить файл... / Добавить файлы...	Скопировать выбранные файлы в исходники
Добавить папку...	Скопировать содержимое папки
Список	Показать текущий состав исходников
Удалить	Убрать одну выбранную позицию
Сбросить	Вернуть исходное состояние

Пробелы и кириллица в путях поддерживаются.

Проверяйте список перед запуском. Команда работает с подготовленным набором файлов, а не с последним, который вы открывали в проводнике.

3. Как устроен экран команды

Каждая команда открывается одинаково.

- Сверху — название команды, кнопка **НАЗАД** и кнопка **ЗАПУСТИТЬ**.

- **Параметры** — поля, сгруппированные по смыслу: Файлы, Листы, Колонки, Строки данных, Территория, Опции.
- **Дополнительно** — поля, которые обычно не нужны: пути к папкам данных, ручные переопределения.
- **Журнал операции** — вывод в реальном времени. Операцию можно отменить.
- **Артефакты** — три кнопки после завершения: **OUTPUT**, **REPORT**, **ПРОВЕРКА**. Открывают папку с результатом, отчёт и аудит-лист.

Одновременно выполняется одна операция. Пока идёт прогон, вторая команда не запустится.

Операции, которые изменяют управляемую рабочую область, спрашивают подтверждение и показывают, что именно будет затронуто.

Как задаются колонки

Буквами, как в Excel: **B**, **C, D**, **AC**. Порядковые номера тоже принимаются. Несколько колонок — через запятую.

Пустое поле почти всегда означает автоматический режим, а не ошибку.

Там, где есть поле «заголовок», колонку можно выбрать по названию из первой строки книги. Явно заданная буква имеет приоритет над заголовком.

4. Команды

Плитки идут в этом порядке:

1. ОКТМО / ключи проекта
2. Сбор адресов
3. Сопоставление адресов
4. Нормализация эталона
5. Обработка эталона
6. Сравнить две таблицы
7. Перенести колонки по адресу
8. Конвертация DOC/XLS
9. Диагностика

4.1. ОКТМО / ключи проекта

Реестр территорий, ключи поиска и пины. Реестр **скачивается кнопкой** и обновляется той же кнопкой, кэш перестраивается сразу.

Это отдельный документ: [ОКТМО и ключи проекта](#).

4.2. Сбор адресов

Рекурсивно собирает адресоподобные значения из XLSX, DOCX, ODT, TXT, CSV, MD и PDF с текстовым слоем, объединяет свидетельства по слотам, отбрасывает неполные строки и пишет книгу адресов.

Применяется, когда сначала нужна инвентаризация, а не сопоставление.

Действие

- *Создать новую таблицу* — собрать всё с нуля.
- *Добавить адреса к готовой* — дописать собранное в существующую книгу.

Источники. «Папка источников» — необязательная папка для рекурсивного обхода; пусто означает всю папку исходников. «Папка адресной таблицы» — где лежит готовая книга адресов.

Колонки. «Адресные колонки» ограничивают разбор конкретными колонками XLSX; пусто — сканируются все ячейки и все поддерживаемые файлы. Для режима дописывания задаются «Колонка собранного адреса» и «Целевая адресная колонка».

Слоты. «Включённые слоты» определяют, из чего собирается адресная строка; «Порядок слотов» — в каком порядке. Поле «Адресная строка» показывает пример получающегося адреса, менять его нельзя. Полный список слотов — в [Справочнике](#).

Опции

Опция	Что делает
Пропускать микрорайон при наличии улицы	Колонка микрорайона остаётся, но в собранную строку не входит
Чистая новая книга	Компактная книга только с адресной колонкой вместо полной
Схлопывать дубли	Объединить кандидатов с одинаковым адресным ключом и выбрать лучшие значения слотов голосованием

Опция	Что делает
Чистить неполные адреса	Убрать улицу без дома, дом без улицы, адрес из одного региона и табличный мусор
Писать лист отбраковки	Отбракованные кандидаты уходят на отдельный лист для аудита
Включить ключи ОКТМО	Определять территорию по реестру и ключам проекта

Эталон для сверки. Необязательная книга с контрольной адресной колонкой. Используется только для оценки качества сбора и пишет `report/address_collection_benchmark.json`.

Результат. Книга `AddressCollection*.xlsx` в папке результата и отчёт `report/address_collection_summary.json`.

После сбора проверьте: пустые значения, дубли, ячейки с несколькими адресами и строки, где есть организация, но нет адреса.

4.3. Сопоставление адресов

Основной режим. Сравнивает адресные колонки с эталонной, переставляет строки так, чтобы адреса встали напротив своих пар, и переносит связанные колонки вместе с ними.

Колонки. «Колонка GT» — эталонная адресная колонка (пусто = автоопределение). «Адресные колонки» — что сопоставлять; пусто означает взять до пяти колонок справа от эталона.

Режим выравнивания

- *Быстрое выравнивание* — сопоставление без дополнительной обработки.
- *Выравнивание + нормализация* — тот же поиск, затем очистка адресных колонок результата.

Связанные колонки. Значения из связанных колонок едут вместе с найденным адресом, чтобы данные не разъехались.

Режим	Поведение
Авто	Прежняя автоматика
Нет	Чистое сопоставление адресных колонок
Слева	Колонки левее каждого адреса
Справа	Колонки правее каждого адреса

Режим	Поведение
Всё между GT и адресом	Одна дальняя колонка тащит за собой весь блок между эталоном и собой

Последний режим отвечает на частый случай: эталон в **A**, адрес в **AC**, и всё **B:AB** должно переехать вместе с **AC**.

Поле «Сателлитные колонки» задаёт связи явно, если автоматика не подходит.

Выравнивание и нормализация. «Нормализовать до сопоставления» — более медленный режим с общим нормализатором. «Собрать до сопоставления» строит чистых кандидатов из выбранных колонок перед поиском; «Весь документ» расширяет сбор на всю книгу. «Город по умолчанию» задаёт название города, которое вычищается из ключей при нормализованном сопоставлении.

Результат. Книга результата в папке вывода и **report/address_alignment_summary.json**. В журнале — сводка по каждой колонке: сколько строк нашлось каждым проходом, сколько точных, сколько мягких, сколько осталось без пары.

Как читать сводку. Высокий процент мягких попаданий — повод посмотреть глазами. Точное совпадение доказательно, приблизительное — гипотеза.

4.4. Нормализация эталона

Создаёт эталонный справочник адресов из источников или приводит существующий к текущему набору слотов.

Действие

- *Сгенерировать* — новый справочник из источников.
- *Обновить* — дополнить существующий.
- *Чистка слотов* — перестроить устаревшие слот-колонки.

Опции. «Схлопывать старые дубли» объединяет записи с одинаковым ключом слотов.

«Перестроить старые слоты» удаляет устаревшие слот-колонки и пишет только текущий набор.

«Чистая новая книга» даёт компактный результат.

Результат. Книга **AddressReference_*.xlsx** и **report/reference_normalization_summary.json**.

Нормализация детерминирована: пробелы, пунктуация, регистр, сокращения и выделение составляющих обрабатываются одинаково от прогона к прогону. Исходное значение адреса

сохраняется в результате.

Нормализованный ключ помогает поиску, но не заменяет доказательную исходную строку.

4.5. Обработка эталона

Сортирует эталонную адресную колонку по слотам или раскладывает её на отдельные слот-колонки.

Действие: *Сортировать, Разложить, Сортировать + разложить.*

Порядок слотов здесь определяет иерархию сортировки: какой слот доминирует первым. Регион → муниципалитет → населённый пункт → улица → дом, или любой другой порядок.

Заголовки слот-колонок: *Полные* (`Postal_Index` , `Municipality`) или *Короткие* (`Index` , `Municipality`).

Удалить исходную адресную колонку — при разложении слот-колонки вставляются, а исходная убирается.

Результат. Обработанная книга и `report/reference_processing_summary.json`.

4.6. Сравнить две таблицы

Берёт любые две книги, собирает адрес из указанных колонок каждой, нормализует, сопоставляет дома и создаёт **отдельную** книгу сравнения. Обе исходные книги остаются нетронутыми.

Поля. Таблица А и Таблица В, листы, адресные колонки каждой, первая строка данных каждой. Адресные колонки В могут стоять на других позициях, чем в А.

Сортировать адреса внутри групп. Выключено — сначала идут сопоставленные строки в порядке А, затем «только А» и «только В». Включено — внутри каждой группы строки отсортированы по адресу.

Результат. Книга сравнения со статусной колонкой и цветовой заливкой:

Статус	Цвет	Значение
сопоставлено	зелёный	Ровно одна пара
сопоставлено: N строк В	жёлтый	Несколько кандидатов, нужен просмотр
только А	оттенок А	Пары нет

Статус	Цвет	Значение
только В	оттенок В	Пары нет

Плюс `report/table_comparison_summary.json`.

4.7. Перенести колонки по адресу

Находит одинаковые дома в двух книгах и переносит выбранные колонки из источника в соответствующие строки основной книги.

Простое копирование строк здесь не годится: порядок строк основной книги, её формулы, оформление и листы должны остаться на месте. Поэтому перенос идёт через Excel, а основная книга сначала копируется в папку результата и правится уже копия.

Поля. Основная книга и книга-источник, листы (в источнике можно отметить несколько), адресные колонки обеих, «Поля для переноса» — что именно переносить, первые строки данных и строка заголовка результата.

Результат. Копия основной книги с добавленными колонками, статусной колонкой и листом аудита `AUDIT_ADDRESS_JOIN`. Статусы: `сопоставлено`, `сопоставлено: N записи`, `не найдено`. Отчёт — `report/safe_table_join_summary.json`.

Проверьте, что один адрес не разрешился в несколько несовместимых строк: именно это показывает статус с числом.

4.8. Конвертация DOC/XLS

Превращает старые `.doc` и `.xls` в исходниках в `.docx` и `.xlsx`.

Перезаписывать существующие — выключено, существующие `.docx`/`.xlsx` сохраняются. Дополнительно удаляются одноимённые PDF/TXT/CSV/MD дубликаты, оставшиеся от прежних конвертаций.

Отчёт — `report/legacy_office_conversion_report.json`.

4.9. Диагностика

Две подкоманды.

Диагностика парсера извлекает разобранные составляющие адреса из исходников и показывает их как есть. Это способ понять, почему конкретный адрес не нашёл пару: видно, что программа сочла улицей, что домом, а что территорией.

Проверить исходники перечисляет поддерживаемые таблицы в папке исходников и попутно убирает временные файлы Office.

5. Обслуживание

Команда	Что делает
Очистить папки	Удалить содержимое управляемых <code>input</code> и <code>output</code> , сами папки остаются
Очистить пустые строки	Создать копии результатов без строк, пустых по всем колонкам
Упаковать результаты	Собрать <code>release/address_aligner_output.zip</code> из папки результата

Очистка пустых строк убирает только строки, пустые целиком. Строка с данными в одной-единственной колонке считается непустой и остаётся.

Очистка папок ограничена управляемой рабочей областью: `cleanup_managed_workspace_only` и `never_delete_original_inputs` в `config/project.yaml` не дают ей выйти за пределы проекта и удалить оригиналы.

6. Порядок работы

Типовой прогон:

1. **Подготовьте реестр ОКТМО**, если работаете с территориальным контекстом — один раз кнопкой, дальше по мере выхода новых срезов.
2. **Наберите ключи** региона или муниципалитета на экране ОКТМО.
3. **Сложите исходники**, проверьте список.
4. **Прогоните на маленьком контролируемом наборе** — десяток строк, где ответ известен заранее.
5. **Запустите полный прогон.**
6. **Прочитайте отчёт.** Прогон не закончен, пока отчёт не просмотрен.
7. **Откройте результат** в табличном приложении и сверьте заголовки, порядок строк, формулы и перенесённые значения.

Что смотреть в отчёте

- пустые адреса;
- дубли ключей;
- связи «один ко многим» и «многие к одному»;
- расхождения по дому и корпусу;
- кандидаты с низкой уверенностью;
- строки, не попавшие в результат;
- неожиданное изменение количества строк.

Правила безопасной работы

Сохраняйте исходный порядок строк, если он важен для систем, которые заберут результат дальше.

Не нормализуйте руками часть набора перед автоматическим сравнением: половина приведённых к одному виду строк и половина сырых дают худший результат, чем целиком сырой набор.

Держите исходники и отчёт до приёмки результата.

После изменения логики сопоставления сначала прогоняйте маленький контролируемый набор.

При сбое сохраните журнал и отчёт. Исправьте источник, эталон или разметку колонок — и повторяйте прогон только после того, как поняли причину.