

Руководство пользователя

**программа для ЭВМ
«Искусственный интеллект «АТОМ»**



Правовая информация

Авторское право © 2022, ЗАО «Сервис-Газификация». Все права защищены.

Функциональные возможности и пользовательский интерфейс программы для ЭВМ «Искусственный интеллект «АТОМ» (далее по тексту равнозначны полному наименованию: «ИИ «АТОМ»», «ИИ», «приложение «АТОМ»), описанные в настоящем руководстве, могут быть изменены в последующих версиях программного продукта без уведомления пользователей.

ИИ «АТОМ», описанный в настоящем документе, передается по лицензионному соглашению и может использоваться или копироваться только в соответствии с полученной лицензией.

ЗАО «Сервис-Газификация» предоставляет пользователям, обладающим лицензией, право на печать копий настоящего документа или его частей исключительно для личного использования.

Содержание

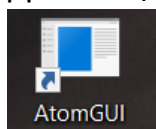
1	О приложении «АТОМ»	2
2	Работа с документами в приложении «АТОМ»	3
2.1	Открытие документа/папки в приложении	3
2.2	Выбор типа документа	6
2.3	Предварительный просмотр	10
2.4	Сохранение результата обработки данных	12

1 О программе для ЭВМ «Искусственный интеллект «АТОМ»

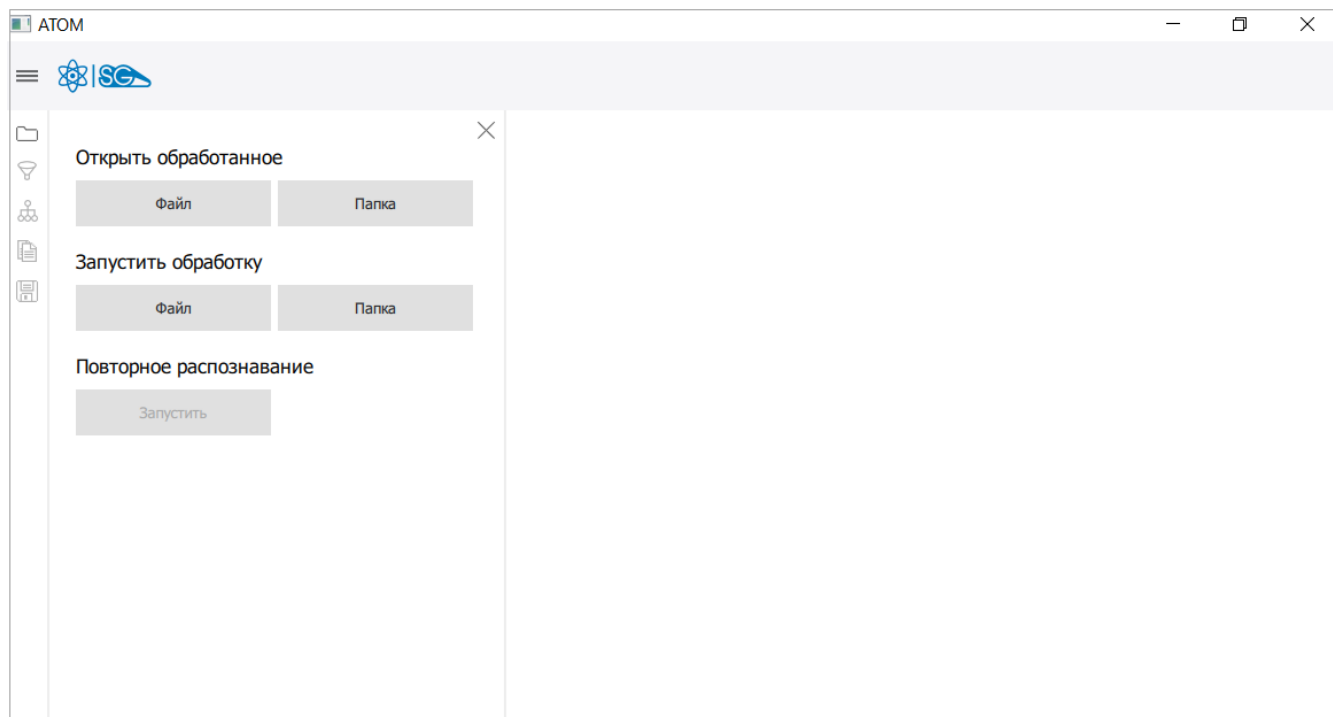
Программа для ЭВМ «Искусственный интеллект «АТОМ» предназначена для работы со сканированными изображениями печатных документов в формате PDF с целью извлечения из них определенных текстовых данных (необходимых для решения конкретных задач) и сохранения их в редактируемом файле .csv.

При разработке приложения «АТОМ» использованы технологии интеллектуальной обработки документов, в частности **OCR** (Optical Characters Recognition) – технология оптического распознавания символов и их извлечения из сканов и PDF-файлов.

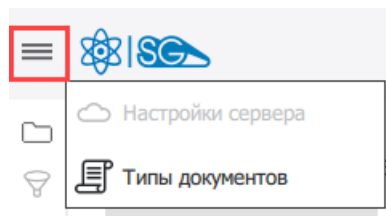
Для входа в приложение **АТОМ** дважды щелкните мышью по его значку



на рабочем столе. Откроется *главная страница* приложения.

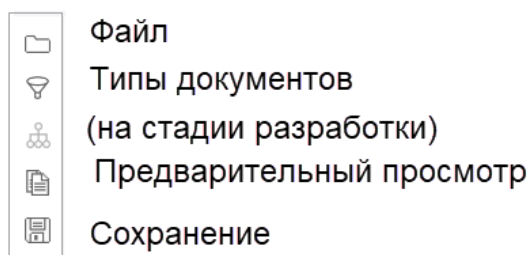


ИИ «**АТОМ**» снабжено компактным и интуитивно понятным пользовательским интерфейсом. В верхнем левом углу главной страницы находится меню настроек, открывающееся нажатием кнопки



Данный функционал используется для настройки сервера и параметров приложения и предназначен для системных администраторов.

Под кнопкой меню настроек расположена боковая панель инструментов, каждая кнопка которой открывает соответствующий раздел приложения:




Далее приводится подробное описание всех разделов.


2 Работа с документами в ИИ «АТОМ»

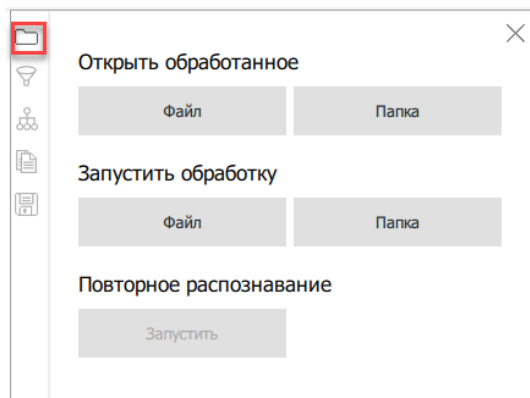
Процесс обработки сканированных печатных документов, включая сканы чертежей и документов с табличными данными, состоит из нескольких этапов:

- Открытие документа или папки с документами формата **PDF** в приложении **АТОМ**
- Выбор типа документа исходного скана для извлечения из него необходимых текстовых данных
- Предпросмотр и редактирование (при необходимости) результата обработки скана
- Сохранение выборки извлеченных данных в редактируемом файле текстового формата **.csv** и обработанного документа в формате **PDF**



2.1 Открытие документа/папки в приложении

Кнопка  на панели инструментов приложения **АТОМ** используется для открытия раздела **Файл**. При запуске приложения данный раздел

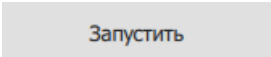
открывается по умолчанию. Чтобы закрыть его, нажмите  в правом верхнем углу раздела.



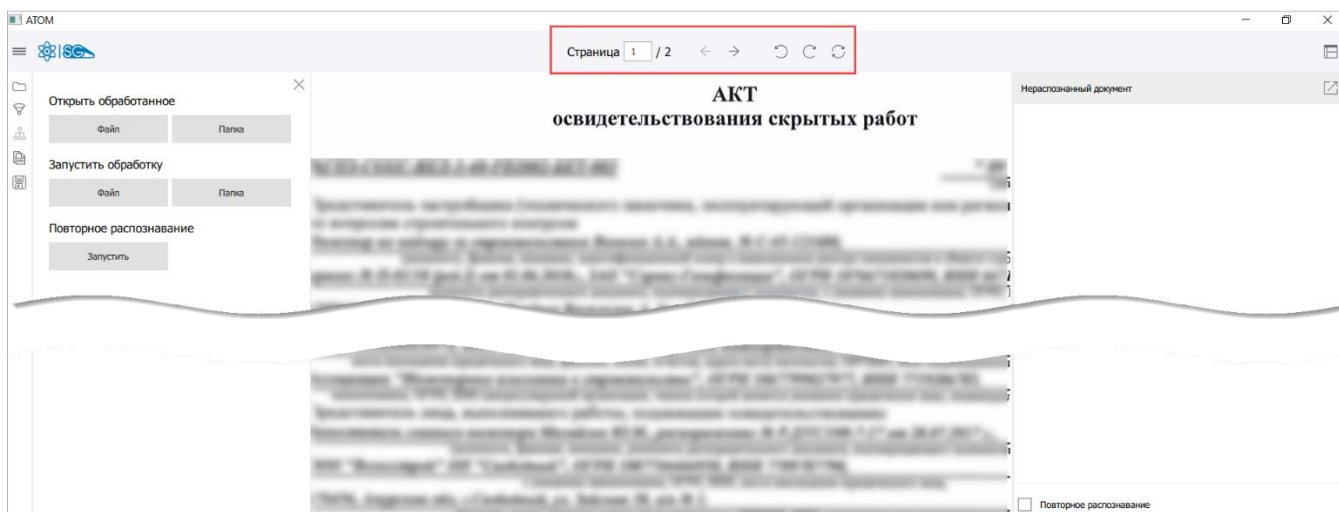
*В приложении **АТОМ** предусмотрена работа только со сканированными документами в формате **PDF***

Чтобы открыть файл формата **PDF** либо папку с **PDF**-файлами, нажмите кнопку  или  соответственно. При этом открывается окно **Выберите файл** или **Выберите папку**, в котором необходимо произвести выбор нужного документа или папки.

Нажатие кнопок **Файл** и **Папка** в пункте **Открыть обработанные** открывает уже обработанные в приложении файлы документов; нажатие данных кнопок в пункте **Запустить обработку** загружает в приложение сканы документов, предназначенные для обработки.

Чтобы выполнить повторную обработку документа, нажмите кнопку  в пункте **Повторное распознавание**.

Открытый документ или последовательно расположенные документы из папки отображаются в центральной части приложения.



Непосредственно над окном документа располагается панель навигации:



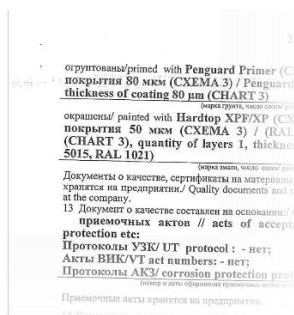
Открыть нужную страницу можно путем пролистывания документа колесом мыши. При этом в поле указывается номер текущей открытой страницы. Ввод в данное поле нужного номера страницы и нажатие клавиши **Enter** также позволяет открыть требуемую страницу документа.


Нажмите кнопку , чтобы вернуться на предыдущую страницу, а кнопку – для перехода на следующую.

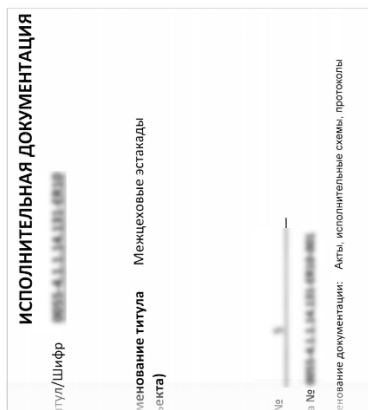
Правильная ориентация сканов необходима для корректного оптического распознавания символов системой. Неверную ориентацию можно исправить с помощью кнопок корректировки угла поворота скана





. Кнопки и применяются для поворота страниц на небольшой угол (до 2°) влево и вправо соответственно. Например, данный скан нужно слегка повернуть влево, чтобы текст документа располагался точно по горизонтали:

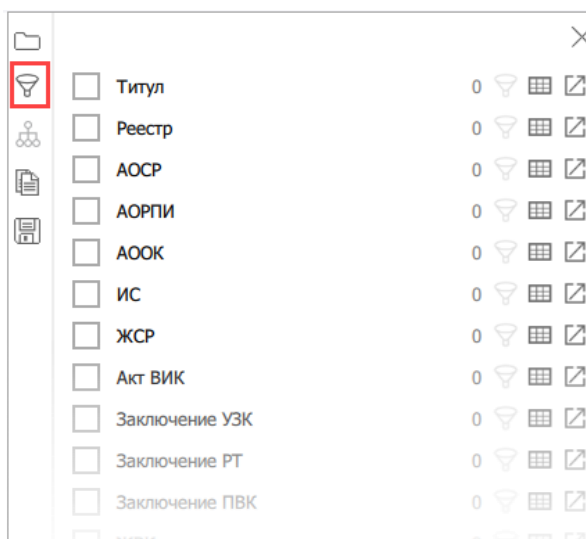


Для поворота скана документа на угол 90° по часовой стрелке нажмите кнопку :



2.2 Выбор типа документа

Кнопка  на панели инструментов ИИ «АТОМ» используется для открытия раздела **Типы документов**. Чтобы закрыть его, нажмите  в правом верхнем углу раздела.



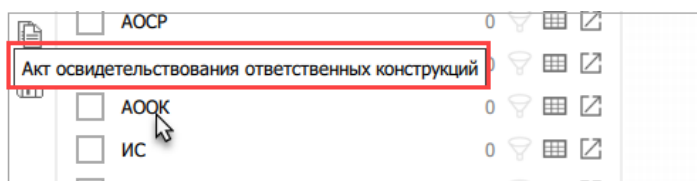
В разделе **Типы документов** представлен список типизируемых документов.



Данный список является предустановленным в приложении. Для добавления в него нового типа документа требуются специальные права, которыми обладает системный администратор. Обратитесь к нему, если возникнет необходимость пополнения списка


Для каждого типа документов система распознает и извлекает из скана определенную выборку данных, соответствующую конкретному типу. Например, для типа документа **Спецификация** распознаются и извлекаются следующие данные:


Выберите тип документа, щелкнув мышью по его строке в списке. Для удобства выбора типа воспользуйтесь всплывающей подсказкой, которая открывается при наведении на него курсора:



В строке выбранного типа документа устанавливается синяя метка , и указывается количество документов в обработанной папке:



Кнопка  предназначена для фильтрации по типу документа. При пролистывании колесом мыши или кнопками навигации открываются только документы, для типа которых включен фильтр. Можно установить сразу несколько фильтров для разных типов документов.

Кроме того, по каждому типу документа можно открыть сводный реестр документов, нажав кнопку .

	Объект строительства	Застройка	ГСП	Лицо осуществляющее подготовку...	Представитель заказчика	Представитель ГСП	Проект
1							
2							





В сводном реестре в виде таблицы содержатся все извлеченные данные по выбранному типу документа. Щелкнув мышью по ячейке таблицы, можно перейти на соответствующий исходный скан документа, из которого извлечены данные.


(Кнопка  в строке типа документа находится на стадии разработки).


Смена типа осуществляется только для текущей страницы документа. Если это не первая страница, то предыдущие страницы сохраняют свой тип.



Все последующие страницы (при их наличии) помечаются как *нераспознанные* документы, которые можно присоединить.

Справа от окна скана находится **Область данных документа**, в которой отображается результат обработки скана, содержащий поля с текстовыми данными, извлеченными для выбранного типа документа.

Спецификация 			
Общая часть	Таблица 1	Таблица 2	Таблица 3
№ спецификации			
<input type="text" value="12345"/>			
Подпись			
<input type="text" value="да"/>			
Руководитель работ			
<input type="text" value="Иванов И.И."/>			

В данной области можно просматривать и редактировать распознанные поля и таблицы документа. Каждая таблица находится в отдельной вкладке, между которыми можно переключаться. Область можно открыть в отдельном окне путем нажатия кнопки  в правом верхнем углу.

Кнопка , расположенная над областью данных документа, служит для переключения режимов ее отображения с вертикального на горизонтальный.

Чтобы отредактировать текст, нажмите кнопку  в соответствующей строке. Измените в поле текст и для подтверждения нажмите .

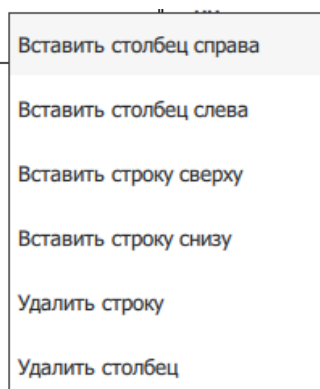
№ спецификации
12345
<input checked="" type="checkbox"/>

Для редактирования полей таблицы щелкните левой кнопкой мыши по нужной ячейке и введите текст. Завершить редактирование можно клавишей **Enter**.

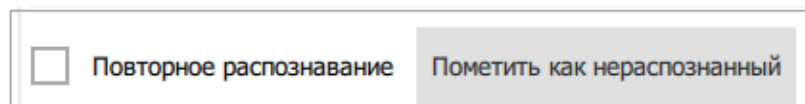
При выборе мышью ячейки производится переход на ту страницу, где была распознана данная ячейка.

Общая часть	Таблица 1	Таблица 2	Таблица 3
1	2	3	4
Наименование	DN, мм	Материал, НД	Количество, м
2	3	4	5
Труба LTCS A671 CCKO CL22 BE EPW 13.0599H WT, Item code: 5021819	813,419,05	A671 CCKO CL22	1,7
Труба LTCS A333-6 BE 9MLS STD WT, Item code: 5021106	114,346,02	ASTM A333, Gr.6	3,3
Труба LTCS A333-6 BE 9MLS HS, Item code: 5021105	80,345,54	ASTM A333, Gr.6	0,3

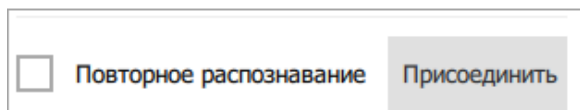
Щелчок правой кнопкой мыши по ячейке открывает контекстное меню, с помощью которого можно добавлять и удалять столбцы и строки до и после выбранной ячейки, а также удалять строку и столбец, в которых находится выбранная ячейка.



В нижней части области находится опция **Повторное распознавание** и кнопка **Пометить как нераспознанный** (либо кнопка **Присоединить**).





или

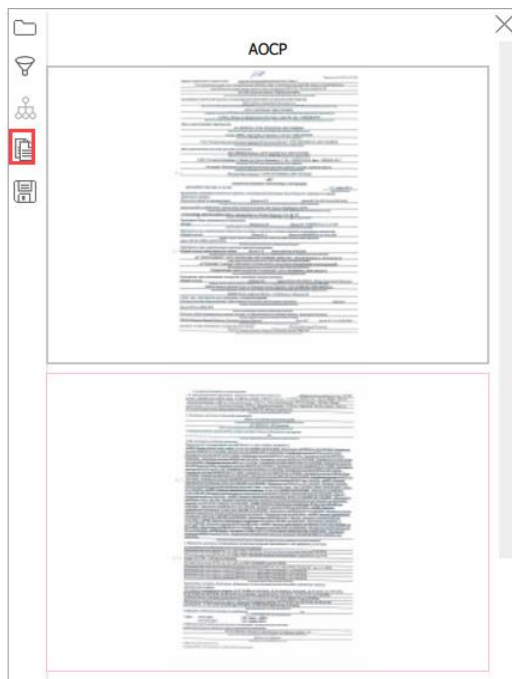


Функция **Повторное распознавание** удобна в случаях, когда у документа были недостающие страницы, позднее добавленные вручную, либо ошибки определения ориентации страниц. Все отмеченные документы распознаются повторно после нажатия кнопки **Запустить** в пункте **Повторное распознавание** раздела **Файл**.

Кнопки **Пометить как нераспознанный** и **Присоединить** являются аналогами опций контекстного меню, открывающегося щелчком правой кнопкой мыши по миниатюре страницы в разделе **Предварительный просмотр**.

2.3 Предварительный просмотр

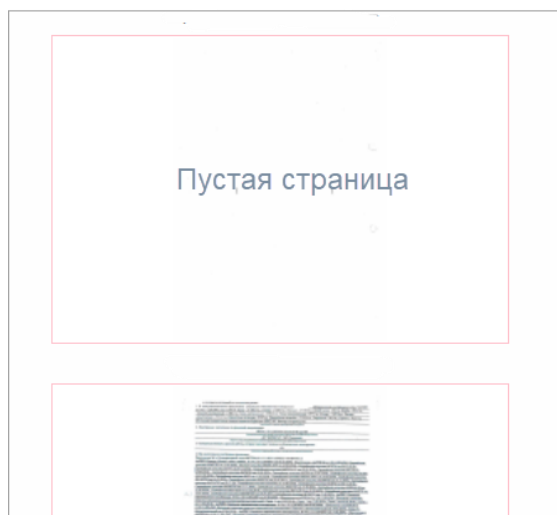
Кнопка  на панели инструментов ИИ «АТОМ» используется для открытия раздела **Предварительный просмотр**. Чтобы закрыть его, нажмите  в правом верхнем углу раздела.



В окне **Предварительный просмотр** отображаются миниатюры всех страниц документа. Чтобы открыть выбранную страницу, щелкните по

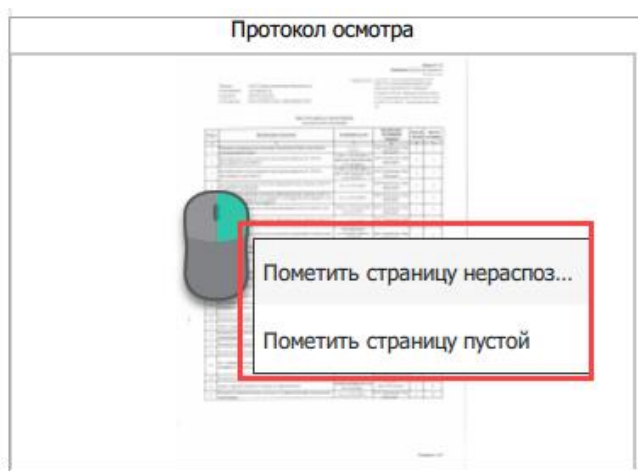
нужной миниатюре левой кнопкой мыши. Над первой страницей указывается наименование типа документа, а на последующих страницах содержится все, что к нему относится.

В документе могут также присутствовать страницы, не содержащие текстовых данных. Система распознает их как *пустые*, а в окне предпросмотра они имеют отметку **Пустая страница**.



Никакие действия с пустыми страницами не производятся, и в итоговом документе они отсутствуют.

Смена статуса страницы осуществляется в контекстном меню, которое открывается щелчком правой кнопки мыши по миниатюре страницы.

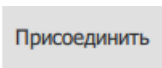


В настоящее время статус можно менять только у текущей открытой страницы.



Любую страницу можно пометить пустой либо снять с нее этот статус. Если документ не является нераспознанным, то любая страница в нем может быть помечена как *нераспознанная* (данная опция удобна в случае, когда

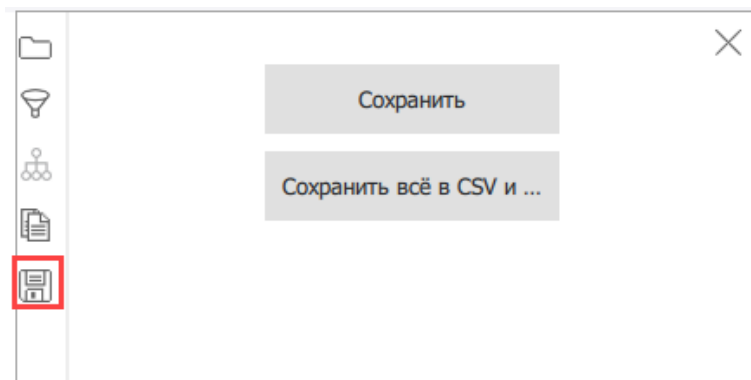
была захвачена лишняя страница, которая не должна быть типизирована). При этом нераспознанными становятся все страницы до конца документа. Любую нераспознанную страницу можно присоединить к ближайшему типизированному документу (при его наличии), и все нераспознанные страницы, предшествующие присоединяемой, также добавляются в этот документ.

Если многостраничный документ не распознан и все его страницы имеют отметку **Нераспознанный документ**, необходимо выполнить повторное распознавание такого документа:

- В окне предпросмотра щелкните мышью по первой странице документа и в разделе **Типы документов** выберите нужный тип
- В окне предпросмотра щелкните мышью по последней странице документа и выберите в контекстном меню опцию **Присоединить** (либо нажмите кнопку  внизу области данных документа)
- Установите флажок рядом с опцией **Повторное распознавание** внизу области данных документа
- Запустите повторное распознавание документа в разделе **Файл**, нажав кнопку **Запустить** в пункте **Повторное распознавание**.

2.4 Сохранение результата обработки данных

Кнопка  на панели инструментов приложения **АТОМ** используется для открытия раздела **Сохранение**. Чтобы закрыть его, нажмите  в правом верхнем углу раздела.



Кнопка **Сохранить** сохраняет текущие изменения и позволяет позднее вернуться к редактированию. При нажатии кнопки **Сохранить в CSV и PDF** создаются реестры **.csv** и осуществляется сохранение нарезанных pdf-файлов.

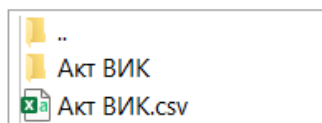
Просмотреть итоговые документы можно в каталоге:

АТОМ\data\export\название родительской папки\название обрабатываемого файла или папки.

Например, для исходного скана с именем **Акт ВИК**, находящегося в папке **Трубопроводы**, путь к папке с результатами обработки будет выглядеть так:

АТОМ\data\export\Трубопроводы\Акт ВИК.

В ней содержится вложенная папка с файлами **PDF** и сводный реестр **.csv** для каждого типа документов. Папка с PDF-файлами и реестр **.csv** имеют одинаковые имена:



Результирующие файлы **PDF** разбиты по типам документов, в них отсутствуют пустые страницы и скорректирована ориентация страниц.

Сводный реестр формата **.csv** имеет следующий вид:

Номер закл	Дата	Наименов	Титул	Номер ли	Номер че	Категория	Заклучен	Лаборато	Файл	Оригинал	Кол-во ст	Страница	Страница общ
0032-МК-	от/issued3	НаиМено	4.1.1.14.1	0-91-WPC	SR09101-0-91-WPC	#ИМЯ?	ООО "АКС	Акт ВИК_2	Акт ВИК_ц		1	1	24

В нем представлены извлеченные данные по всем документам определённого типа. Кроме того, в столбце **Файл** указывается имя результирующего PDF-файла, в столбце **Оригинал** – имя исходного файла сканированного документа, а в последних трех столбцах содержатся сведения о количестве и номере обработанных страниц и общем числе страниц в документе.