

Оглавление

Шаблонизатор Sherpa RPA	2
1.Интерфейс Шаблонизатора	3
1.1 Основное меню Шаблонизатора	4
1.1.1 Меню «Фаѝл»	6
1.1.2 Меню «Обработка»	7
1.1.3 Открыть исходный файл	9
1.1.4 Обработать документ	10
1.1.5 Проверить	11
1.1.6 Добавить атрибут	12
1.1.7 Выполнить полное сканирование	13
1.1.8 Умный помощник	15
1.1.9 Масштаб	17
1.1.10 Меню «Отображать»	17
1.2 Центральная рабочая область Шаблонизатора	18
1.3 Панель настроек	19
1.3.1 Настройки	20
1.3.2 Якоря	33
1.3.3 Атрибуты	38
1.3.4 Таблицы	43
1.3.5 Чертежи таблиц	48
2. Действия в Sherpa Designer после создания шаблона	57
2.1 Значения переменных	57
2.2 Работа с многостраничными документами	63
3. Примеры	66
3.1 Создание шаблона на примере документа «Справка о стоимости выполненн работ»	
3.1.1 Предобработка документа	67
3.1.2 Создание шаблона	69
3.1.3 Работа с таблицами	72
3.1.4 Якоря	83
3.1.5 Создание атрибутов	88
3.2 Создание шаблона с фиксированными атрибутами на примере документа «Счет на оплату»	95



Шаблонизатор Sherpa RPA

Шаблонизатор представляет собой инструмент создания шаблонов (визуальный редактор) и является неотъемлемой частью Дизайнера Sherpa RPA. Шаблонизатор не привязан к каким-либо облачным сервисам и входит в комплект поставки.

Основные возможности

- Определение типа документа (текстовый, табличный; одно- или многостраничный и т.д.). В случае, когда роботу требуется создать несколько шаблонов по различным типам документа (например, Акт отгрузки товара, Акт оказания услуг), Шаблонизатор автоматически подбирает наиболее подходящий шаблон из списка;
- Исправление перспективы и перекосов скан-копии (фото) обрабатываемого документа;
 - Зонирование и сегментация документов;
- Автоматический поиск и разметка таблиц, в том числе без видимых границ;
 - Динамическая привязка содержимого;
 - Проверка наличия печатей, подписей, рукописных текстов;
- Извлечение корректной информации из скан-копии/фото документа.

Обработка сложных документов

Процесс обработки сложных документов выглядит следующим образом:

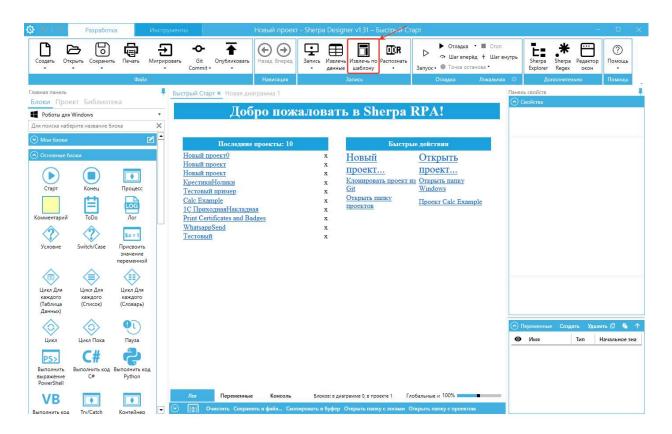
№ п/п	Этап обработки	Описание
1.	Коррекция и	Определение типов документов, разбор по типам и
	типизация	страницам Коррекция перекосов и перспективы
		Зонирование и сегментация



2.	Предобработка Шаблонизатором	Большая часть шаблона создаётся автоматически Автоподбор подходящего шаблона Автоопределение границ таблиц
3.	Обработка OCR	Онлайн или офлайн 5 модулей (движков) на выбор Внешние или встроенные в платформу
4.	Постобработка документа	ML & NLP Man in the loop Проверка наличия подписей, печатей, галочек, рукописных текстов
5.	Занесение в ИС	SAP Oracle R12 1C Битрикс 24 и другие

1. Интерфейс Шаблонизатора

Запуск Шаблонизатора осуществляется на верхней панели основного меню Дизайнера во вкладке «Разработка» — раздел «Запись» — «Извлечь по шаблону».



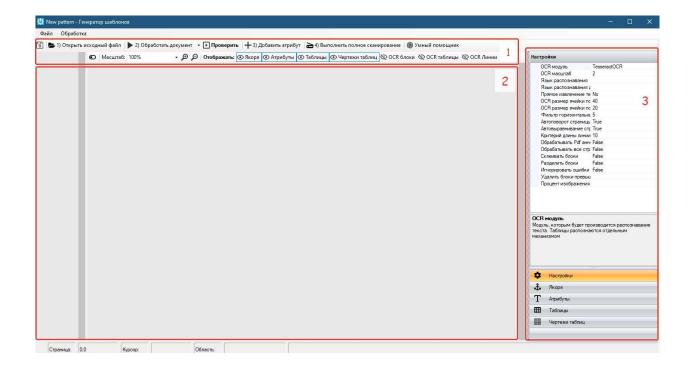
При нажатии на кнопку «Извлечь по шаблону» открывается обработка «Генератор шаблонов» (далее Шаблонизатор).

Пользовательский интерфейс Шаблонизатора состоит из следующих основных областей:

1) Основное меню (верхняя панель);

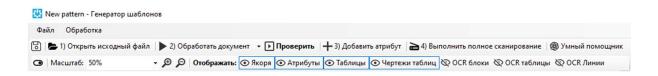


- 2) Центральная рабочая область область для работы с шаблоном обрабатываемого документа;
 - 3) Панель настроек.



1.1 Основное меню Шаблонизатора

Верхняя панель основного меню позволяет увидеть данные, вычисленные встроенным OCR — модулем, а также предлагает основные действия для создания и обработки шаблонов.





Элементы интерфейса основного меню Шаблонизатора.

№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	меню «Файл»	Позволяет использовать базовые функции по управлению файлами документов и шаблонов.
2.	меню «Обработка»	Позволяет производить основные действия по обработке документа.
3.	кнопка «Открыть исходный файл»	Позволяет выбрать и открыть ранее сохраненный сканированный документ, который будет использоваться в качестве образца при создании шаблона для распознавания роботом определенного типа документа.
4.	кнопка «Обработать документ»	Позволяет запустить обработку документа по выбранным параметрам для отображения (Якоря, Атрибуты, Таблицы, Чертежи таблиц, ОСК блоки, ОСК таблицы, ОСК линии).
5.	кнопка «Проверить»	Позволяет визуально обновить макет шаблона после обработки документа и внесения каких-либо правок в шаблон (создание/редактирование якоря/атрибута/таблицы) без повторной отправки документа на обработку ОСК.
6.	кнопка «Добавить атрибут»	Позволяет определить и обозначить области документа, данные которых будут распознаваться роботом и импортироваться в таблицу «Результаты».
7.	кнопка «Выполнить полное сканирование»	Позволяет выполнить полное сканирование.
8.	кнопка «Умный помощник»	Позволяет проанализировать документ и создать шаблон для документа.



9.	переключатель «Масштаб»	Позволяет задать масштаб при работе со сканированным документом.
10.	Меню «Отображать»	Меню «Отображать» позволяет управлять отображением якорей, атрибутов, таблиц и чертежей таблиц, ОСК блоков, ОСК таблиц, ОСК линий в рабочей области.

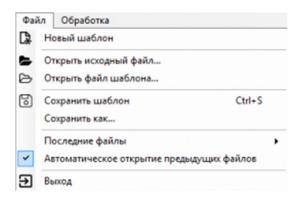
1.1.1 Меню «Файл»

Элементы интерфейса основного меню Шаблонизатора. Меню «Файл».

№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	функция «Новый шаблон»	Позволяет создать новый шаблон.
2.	функция «Открыть исходный файл»	Позволяет выбрать исходный файл для создания шаблона. При нажатии на кнопку открывается окно для поиска исходного файла с возможность указать путь и название файла.
3.	функция «Открыть файл шаблона»	Позволяет выбрать один из ранее созданных шаблон по указанному пути.
4.	функция «Сохранить шаблон»	Позволяет сохранить созданный шаблон по указанному пути. Имя шаблона задается вручную. Также, для сохранения можно воспользоваться сочетанием клавиш Ctrl+S или нажать на в левом верхнем углу панели основного меню. Имя файла нужно указывать, исходя из вида документа, например, «ТОРГ 12». Расширение сохраняемого файла .json
5.	функция «Сохранить как»	Позволяет сохранить созданный шаблон по указанному пути. Имя шаблона задается вручную. Расширение сохраняемого файла .json



6.	функция «Последние файлы»	Позволяет выбрать один из ранее сохраненных файлов из выпадающего списка.
7.	функция «Автоматическое открытие предыдущих файлов»	Позволяет включить автоматического открытие предыдущих файлов. При каждом открытии Шаблонизатора автоматически открывает файл, с которым велась работа до закрытия окна Шаблонизатора. Для активации функции необходимо кликнуть левой клавишей мыши по строке до появления
8.	функция «Выход»	Позволяет выйти из Шаблонизатора.



1.1.2 Меню «Обработка»

Элементы интерфейса основного меню Шаблонизатора. Меню «Обработка».

№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	функция «Обработать документ»	Позволяет выполнить обработку текущего файла по заданным инструментам и атрибутам, выбранным в верхней панели основного меню (Якоря, Атрибуты, Таблицы, Чертежи таблиц, ОСК-блоки, ОСК-таблицы, ОСК-линии). Также, для обработки документа можно воспользоваться сочетанием клавиш Ctrl+R.

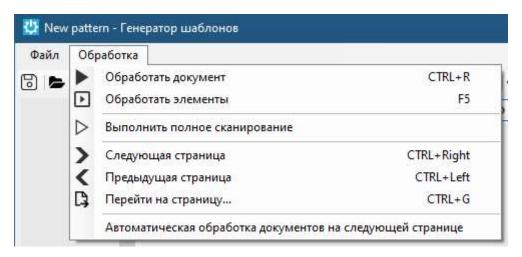


2.	функция «Обработать элементы»	Позволяет выбрать один из ранее созданных шаблонов по указанному пути для его дальнейшей обработки по заданным инструментам и атрибутам, выбранным в верхней панели основного меню. Также, для обработки элементов можно нажать F5.
3.	функция «Выполнить полное сканирование»	Позволяет выполнить полное сканирование. После выполнения действия открывается окно «Результат полного сканирования».
4.	функция «Следующая страница»	Позволяет перейти на следующую страницу в случае, если документ многостраничный. Также, для перехода на следующую страницу можно воспользоваться сочетанием клавиш Ctrl+Right.
5.	функция «Предыдущая страница»	Позволяет перейти на предыдущую страницу в случае, если документ многостраничный. Также, для перехода на следующую страницу можно воспользоваться сочетанием клавиш Ctrl+Left.
6.	функция «Перейти на страницу»	Позволяет перейти на страницу по ее номеру. Также, для перехода на страницу по ее номеру можно воспользоваться сочетанием клавиш Ctrl+G.
7.	функция «Автоматическая обработка документов на следующей странице»	Позволяет выполнить автоматическую обработку следующей открытой страницы по заданным инструментам и атрибутам, выбранным в верхней панели основного меню для текущей страницы. Используется в случае, когда в одном файле находится несколько однотипных одностраничных документов, размещенных на одной странице (например, документ ТОРГ 12). При нажатии кнопки открывается Следующая страница и распознается следующий документ.



При работе с многостраничным документом, содержащим в себе разные типы документации, не рекомендуем использовать эту функцию, т.к. робот будет распознавать все якоря, созданные на каждой странице документа.

Многократный переход между страницами может увеличивать время работы робота.

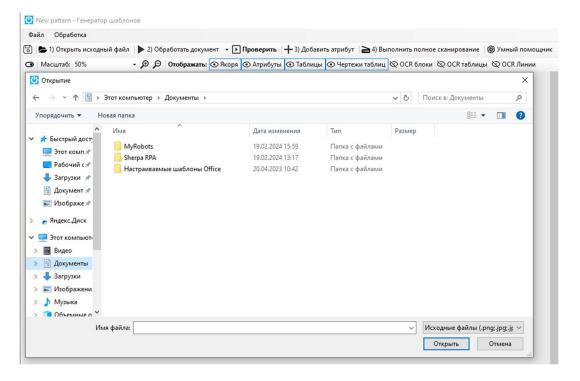


1.1.3 Открыть исходный файл

Шаблонизатор позволяет выбрать и открыть ранее сохраненный сканированный документ, который будет использоваться в качестве образца при создании шаблона для распознавания роботом определенного типа документа.

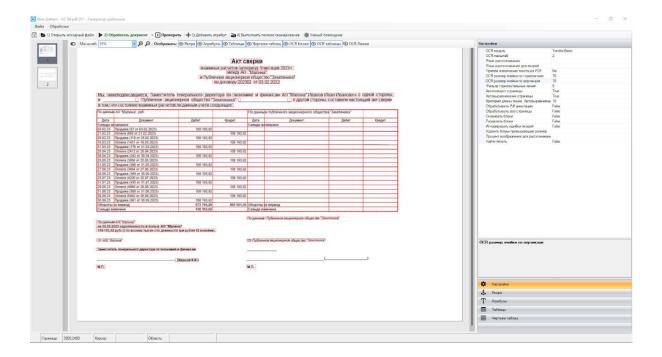
Файл может быть представлен в одном из форматов: .pdf, .png, .jpeg, .jpg. Файл .pdf может быть многостраничным.





1.1.4 Обработать документ

Кнопка «Обработать документ» позволяет запустить обработку документа по выбранным параметрам для отображения якорей, атрибутов, таблиц, чертежей таблиц, ОСR блоков, ОСR таблиц, ОСR линий в рабочей области. Обработанный документ выглядит как показано на скриншоте.

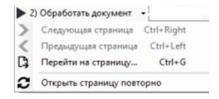




Также, можно выбрать один или несколько критериев на верхней панели основного меню Шаблонизатора.



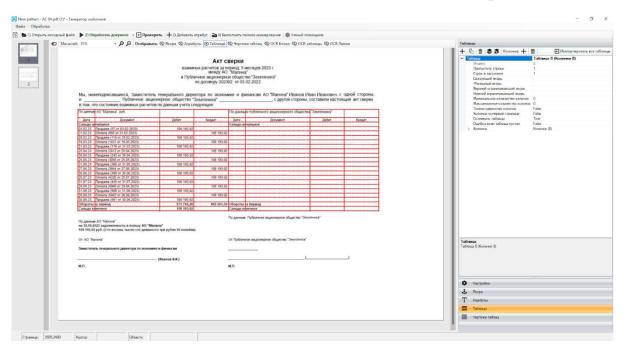
При нажатии на кнопку появляется возможность произвести обработку следующей страницы (Ctrl+Right) или предыдущей страницы (Ctrl+Left), перейти на страницу (Ctrl+G). Кнопка «Открыть страницу повторно» позволяет открыть исходный документ без внесенных изменений в настройках Шаблонизатора. Данная кнопка используется для проверки корректности заданных настроек в Шаблонизаторе. Таким образом можно проверить, например, срабатывает ли настройка «Автоповорот страницы». Если при нажатии кнопки «Открыть страницу повторно» мы видим, что документ изначально был неверно повернут, значит настройки Шаблонизатора были заданы верно и при обработке сработал Автоповорот страницы.



1.1.5 Проверить

Кнопка «Проверить» позволяет проверить корректность отображения табличной части, из которой будут импортироваться данные. Также проверку можно запустить нажатием клавиши F5.



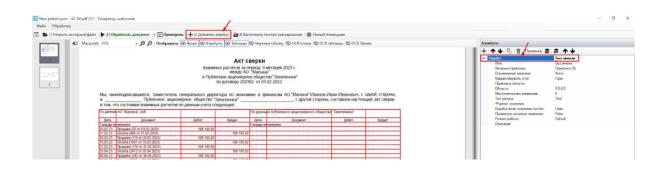


1.1.6 Добавить атрибут

Атрибуты позволяют определить и обозначить области документа, данные из которых будут распознаваться роботом и импортироваться в таблицу «Результаты».

Существует несколько способов добавления Атрибута:

1) Нажатием кнопки на верхней панели Основного меню.



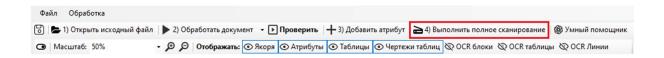
2) Нажатием кнопки + на правой рабочей панели меню во вкладке «Атрибуты».





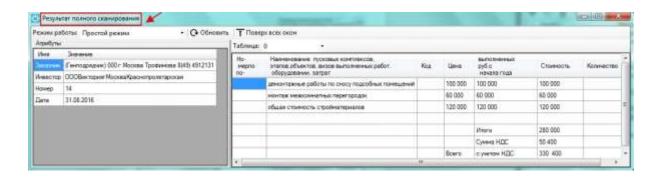
1.1.7 Выполнить полное сканирование

Кнопка «Выполнить полное сканирование» позволяет выполнить полное сканирование документа. После нажатия на кнопку открывается окно «Результат полного сканирования», в котором отображаются распознанные данные из документа в виде заполненной таблицы. Если при проверке обнаружены неточности, то их, в любой момент, можно исправить.



1.1.7.1 Результат полного сканирования

Результат полного сканирования представляет собой окно с результатами, полученными в процессе работы робота со сканированным документом по созданному шаблону. В появившейся форме отображаются все атрибуты и таблицы, обнаруженные в документе.



Окно с результатами поделено на две табличные части:

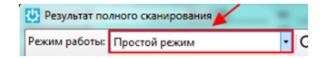
1) Атрибуты (слева) - в данной табличной части отображаются данные, полученные роботом по заданным Атрибутам и Якорям. Чаще всего, это общая информация о Заказчике, Подрядчике, Инвесторе, например, наименование и реквизиты. Также в Атрибутах указываются Номер и Дата документа.



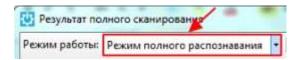
2) Таблица (справа) – в табличной части отображаются данные, содержащиеся в таблицах сканированного документа, полученные роботом в результате работы по созданному шаблону при помощи ОСR-модулей.

Эти данные могут быть числового формата, например, цена, стоимость, сумма и т.д.

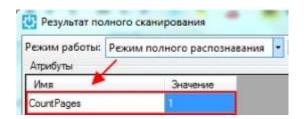
По умолчанию указан режим «Простое сканирование».



При работе с многостраничными документами (если файл содержит в себе два и более документа), на одной из страниц которого указан якорь «LastPage», необходимо использовать «Режим полного распознавания».



При этом будет отображаться выходной параметр – количество страниц, которые участвовали в работе Шаблонизатора.



Полное распознавание будет проводиться с текущей страницы до момента, пока не будет найден якорь «LastPage».

В случае, если мы запускаем «Полное сканирование», находясь не на первой странице многостраничного документа, и, если на одной из страниц указан якорь «LastPage», Шаблонизатор выдаст окно с запросом: «Вы хотите выполнить распознавание с первой страницы?».



При нажатии кнопки «Да» — алгоритм Шаблонизатора будет распознавать документ с первой страницы документа до момента, пока не будет найден якорь «LastPage».

При нажатии кнопки «Нет» — алгоритм Шаблонизатора будет распознавать документ с текущей страницы документа до момента, пока не будет найден якорь «LastPage».



1.1.8 Умный помощник

Кнопка «Умный помощник» позволяет проанализировать документ и создать шаблон для документа.



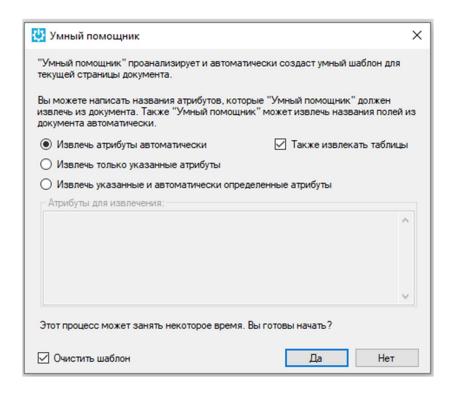
При нажатии на кнопку «Умный помощник» доступны следующие настройки.

Элементы интерфейса основного меню Шаблонизатора. Умный помощник. Настройки.

№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	настройка «Извлечь атрибуты	Позволяет роботу определить значения атрибутов автоматически, без прямого
	автоматически»	указания.



2.	настройка «Извлечь только указанные атрибуты»	Позволяет задать определенные атрибуты для извлечения их из документа (например, «покупатель» или «продавец»).
3.	Настройка «Извлечь указанные и автоматические определенные атрибуты»	Позволяет указать прямо атрибуты, необходимые для извлечения, а также позволить роботу самому определить поля, которые присутствуют в анализируемом документе.
4.	флаг «Также извлекать таблицы»	Позволяет извлечь заголовки таблицы из документа. Если флаг не стоит, то таблица в документе рассматриваться не будет.
5.	флаг «Очистить шаблон»	Позволяет при каждом запуске очистить в шаблоне Якоря, Атрибуты, Чертежи, Чертежи таблиц.



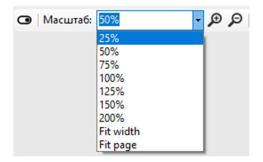
Умный помощник поможет в работе с Шаблонизатором: он покажет, как анализируется документ и создается шаблон.



1.1.9 Масштаб

Переключатель «Масштаб» позволяет задать масштаб при работе со сканированным документом. Размер изображения может быть задан несколькими способами:

- Выбрать значение из выпадающего списка;
- Указать значение вручную;
- Увеличить или уменьшить масштаб при помощи лупы.



1.1.10 Меню «Отображать»

Меню «Отображать» позволяет управлять отображением якорей, атрибутов, таблиц и чертежей таблиц, ОСК блоков, ОСК таблиц, ОСК линий в рабочей области.



Элементы интерфейса основного меню Шаблонизатора. Меню «Отображать».

№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	Якоря	Отображает точки на сканированном
		документе, с помощью которых робот
		определяет, в какой части документа
		размещаются определенные данные.
		Якорей может быть сколь угодно много, в
		зависимости от типа документа.



2.	Атрибуты	Отображает области документа, данные из которых будут распознаваться роботом и импортироваться в таблицу «Результаты».
3.	Таблицы	Отображает границы таблиц, данные из которых будут распознаваться роботом и импортироваться в таблицу «Результаты».
4.	Чертежи таблиц	Отображает чертежи таблиц.
5.	OCR блоки	Отображает блоки, распознанные OCR- модулем.
6.	OCR таблицы	Отображает таблицы, распознанные OCR- модулем.
7.	OCR линии	Отображает линии, распознанные OCR- модулем.

Кнопки Якоря, Атрибуты, Таблицы и Чертежи таблиц относятся напрямую к шаблону. Данные, отображаемые по ним, являются вычисленными после обработки шаблона.



Кнопки ОСR блоки, ОСR таблицы и ОСR линии используются перед обработкой шаблона для настройки шаблона. Они отображают данные, полученные после обработки документа ОСR модулем.



1.2 Центральная рабочая область Шаблонизатора

Область для работы с шаблоном обрабатываемого документа. Здесь отображается исходный файл/сканированный документ, по которому создается шаблон для дальнейшей работы робота.



Dan Dan	остор ООО "Вовгором" з Метов. Едистогр почина (Теппициятия) ООО "Сравног («Кона Уоб (приняская дак тоб) до Станов («Кона Уоб) почина («Кона устано) ООО "Сравного («Кона Уоб) почина («Кона устано») ООО "Сравного («Кона Уоб) почина на почина на по	es, Barril posicia V, 61991 es, Best science, Sci. Cirgo es, Best	7 (1) (1) (4) (8) (1) (1) (1) (4) (8) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1	67)0 67)0 67)0 67)0	54000000000000000000000000000000000000	
No.	Донов Моне ОПРАВБА О СТОИМОСТИ ПАПО ЗИПИНЬ	But a op scapeda op scapeda op scapeda op scapeda op scapeda op scapeda	Вод отера Вод отера	October O	81 04 2018 00 2018 00000000000000000000000000000000000	
po a		Таблиц Строка	а 0 0, Колония 1	- marina	a 1000 raecze ta ornorzadłi sopiany ś	
	Всего работ и заправ, започнативи и - тованить рабея в пол ченице. Вленетильные рабето по смосу подолённого размения месянического поручирация общам изопичесть гарум уразлич общам изопичесть гарум уразлич		100 930 80 000 120 000	100 900 90 500 120 900	100 MB 60 000 1,70 MB	
las.	ина (Геноприяна) теприянці	Kyanana		Course 6.20 Course 6.20 Course 6.20 Xigasana general		
14.1 По Страния 551.666 Курсор:	вален (Субооцияния) — энканатизы	Hamese 5:200:24		House	LA	

При наведении курсора на нижней панели отображаются координаты страницы, курсора и области. Это удобно, когда нужно проверить или скорректировать координаты какого-либо атрибута.

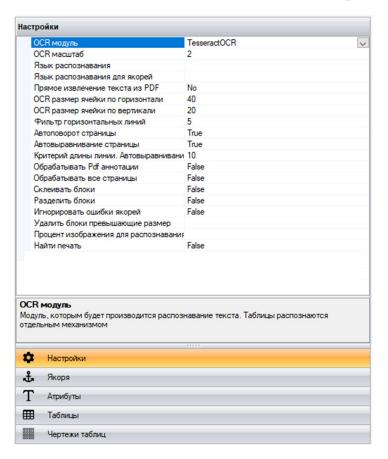
В центральной рабочей области отображаются результаты всех манипуляций, которые мы производим на верхней панели основного меню и на боковой панели настроек Шаблонизатора.

1.3 Панель настроек

В правой панели основного меню Шаблонизатора находятся 5 основных блоков для создания и настройки шаблона:

- Настройки;
- Якоря;
- Атрибуты;
- Таблицы;
- Чертежи таблиц.





1.3.1 Настройки

Шаблонизатор использует несколько видов модулей для извлечения данных из исходных файлов.

OCR модуль

ОСК модуль — модуль, которым будет производиться распознавание текста.

В платформу Sherpa RPA интегрированы несколько ОСR-модулей. Два из них поставляются вместе с роботом: Tesseract OCR и Microsoft OCR. Данные модули могут работать офлайн, без подключения к интернету).

Tesseract OCR — это движок оптического распознавания символов (OCR) с открытым исходным кодом, является самой популярной и качественной OCR-библиотекой. OCR использует нейронные сети для поиска и распознавания текста на изображениях.

Yandex Vision и ABBYY OCR – онлайн-модули, которые используют функции соответствующих облачных сервисов.



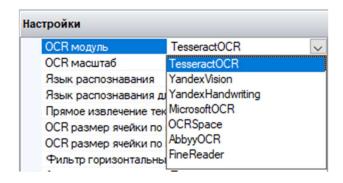
Yandex Handwriting — модуль, позволяющий распознавать рукописный текст.

Microsoft OCR — модуль, позволяющий распознавать текст на изображениях и сканированных документах с помощью оптического распознавания символов (OCR).

OCR Space — модуль, позволяющий распознавать кириллический шрифт на изображениях формата .jpg.

ABBYY FineReader – это коммерческий оффлайн-модуль, использование которого требует отдельной лицензии.

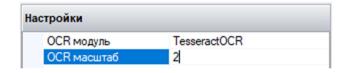
Платформа Sherpa RPA позволяет настроить работу сценария с распознаванием изображений и переключаться между этими ОСR-модулями в любой момент времени.



ОСК масштаб

OCR масштаб — параметр, позволяющий улучшить качество распознавания, если документы имеют плохое разрешение.

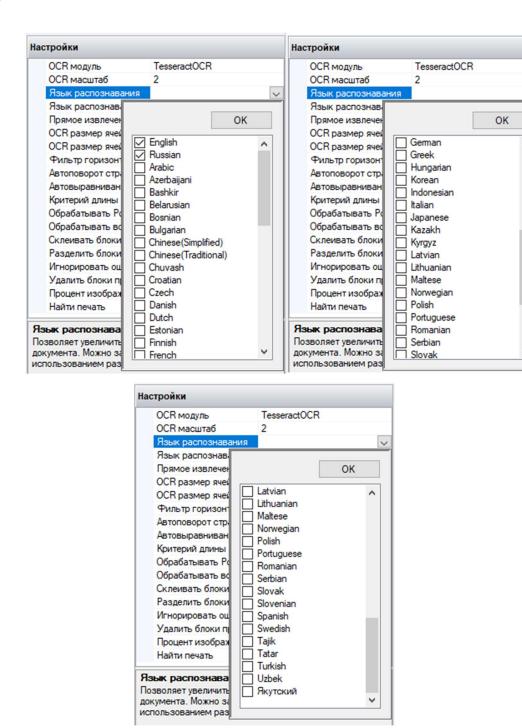
В случае высокого и среднего качества сканированного документа рекомендуется оставить значение масштаба равное «2».





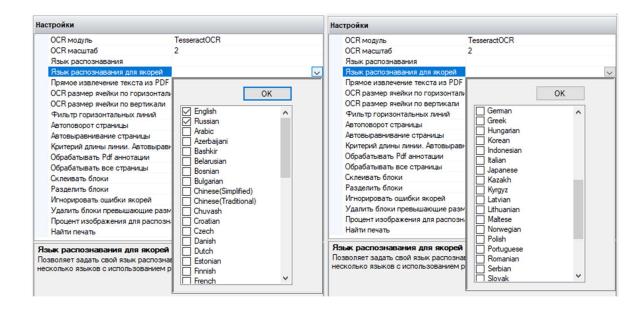
Язык распознавания – параметр, позволяющий увеличить точность распознавания документа. Также, можно задать несколько языков, использовав в качестве разделителя запятые.

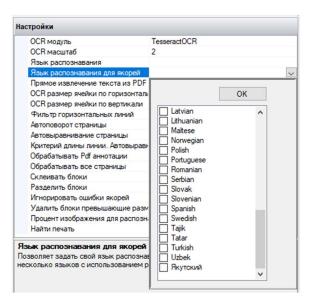
При выборе настройки «Язык распознавания» открывается выпадающий список, в нем с помощью флагов можно отметить нужные для распознавания языки.





Язык распознавания – параметр, позволяющий указывать конкретный язык распознавания для якорей. Также, можно задать несколько языков, использовав в качестве разделителя запятые.





Прямое извлечение текста из PDF

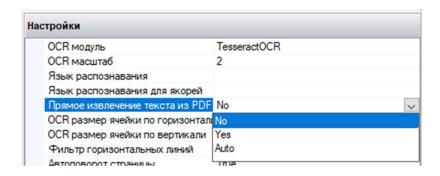
Прямое извлечение текста из PDF – параметр, позволяющий управлять прямым извлечением текста со страницы.

Возможные значения:

- No отключено;
- Yes используется только прямое извлечение текста;



• Auto - автоматический режим (если на странице нет текста, то будет реализовано распознавание текста с заданным ОСК модулем).

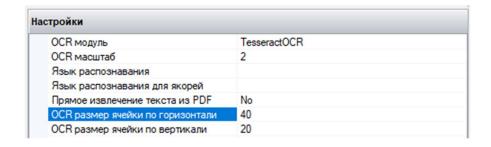


OCR размер ячейки по горизонтали

OCR размер ячейки по горизонтали – параметр, позволяющий указать делитель по горизонтали, по которому определяются ячейки таблиц на странице. Значение должно быть больше или равно 1.

Размер ячейки определяется как размер изображения, деленное на это значение.

Шаблонизатор использует два параметра (по горизонтали и по вертикали) для более точного определения таблиц.



По умолчанию заданы параметры:

40 - OCR размер ячейки по горизонтали

Это значение является оптимальным при распознавании документов со стандартными (или близкими к стандартным) размерами ячеек таблиц.



Рекомендуем оставлять эти значения без изменений и менять их только в случае некорректного распознавания (после проверки).

Если при заданных параметрах (значение по горизонтали 40) таблица не будет распознана, то алгоритм Шаблонизатора увеличит это значение на 10 автоматически и осуществит попытку повторно.

OCR размер ячейки по вертикали

OCR размер ячейки по вертикали – параметр, позволяющий указать делитель по вертикали, по которому определяются ячейки таблиц на странице. Значение должно быть больше или равно 1.

Размер ячейки определяется как размер изображения, деленное на это значение.

Шаблонизатор использует два параметра (по горизонтали и по вертикали) для более точного определения таблиц.

астройк и		
OCR модуль	TesseractOCR	
OCR масштаб	2	
Язык распознавания		
Язык распознавания для якорей		
Прямое извлечение текста из PDF	No	
OCR размер ячейки по горизонтали	40	
OCR размер ячейки по вертикали	20	

По умолчанию заданы параметры:

20 — OCR размер ячейки по вертикали.

Это значение является оптимальным при распознавании документов со стандартными (или близкими к стандартным) размерами ячеек таблиц.

Рекомендуем оставлять эти значения без изменений и менять их только в случае некорректного распознавания (после проверки).



Если при заданных параметрах (значение по вертикали 20) таблица не будет распознана, то алгоритм Шаблонизатора увеличит это значение на 10 автоматически и осуществит попытку повторно.

Фильтр горизонтальных линий

Фильтр горизонтальных линий – параметр, указывающий процент, меньше которого линии будут проигнорированы. Определяется как отношение горизонтальной линии к ширине изображения. Данный параметр используется для атрибутов.

астройки		
OCR модуль	TesseractOCR	
OCR масштаб	2	
Язык распознавания		
Язык распознавания для якорей		
Прямое извлечение текста из PDF	No	
OCR размер ячейки по горизонтали	40	
OCR размер ячейки по вертикали	20	
Фильтр горизонтальных линий	5	
	_	

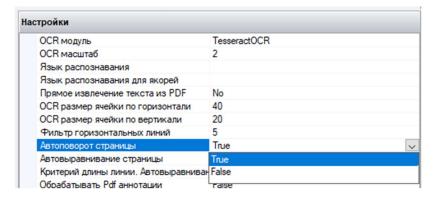
По умолчанию значение данного параметра равно 5.

Часто в сканированных документах присутствуют линии, не являющиеся границами таблиц (различные артефакты, которые необходимо отфильтровать). При заданном параметре все лишние артефакты (линии, нарисованные ручкой или линии, появившиеся при некачественном сканировании) будут проигнорированы.

Автоповорот страницы

Автоматический поворот страницы – параметр, позволяющий автоматически разворачивать страницы на угол, кратный 90 градусам (90, 180 и 270). Автоповорот страницы не разворачивает документ на малые углы. По умолчанию значение параметра «True», рекомендуем оставлять его без изменения.

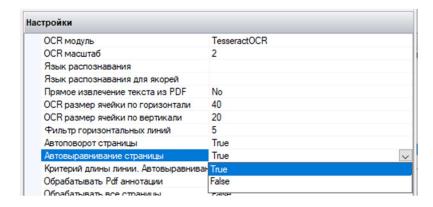






Автовыравнивание страницы

Автоматическое выравнивание страницы – параметр, позволяющий выравнивать содержимое страницы при некорректном сканировании листа. В отличие от Автоповорота страницы, Автовыравнивание разворачивает документ на малые углы.



При автовыравнивании страницы распознается самая длинная ОСR линия, имеющаяся на странице (чаще всего это линия из таблицы или атрибута) и определяет угол ее наклона относительно горизонтали. После этого происходит разворот документа, чтобы найденная линия стала параллельна линии горизонта.

По умолчанию значение параметра «True», рекомендуем оставлять его без изменения.

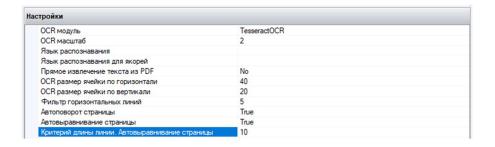
Этот параметр нежелательно использовать в случае, если сканированный документ сильно повернут и угол поворота составляет более 40, в этом случае Шаблонизатор не сможет распознать, в какую сторону нужно выровнять документ и выдаст ошибку. В такой ситуации можно использовать параметр «Корректировка угла».

Критерий длины. Автовыравнивание страницы

Критерий длины. Автовыравнивание страницы – параметр, позволяющий ограничивать длину линии. Данный параметр используется как делитель для ширины страницы. Если длина линии меньше вычисленного значения, то она

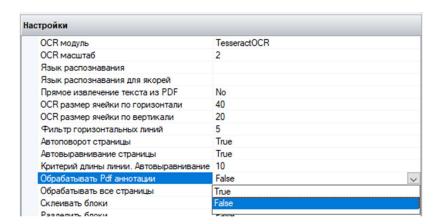


пропускается в алгоритме. Если страница изначально сильно повернута, то следует указать или большее значение или 0 для отключения параметра. По умолчанию значение равно 10.



Обрабатывать PDF аннотации

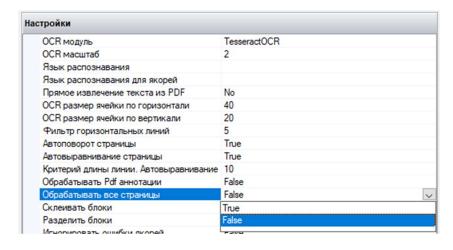
Обрабатывать PDF аннотации – параметр, позволяющий включить обработку аннотаций PDF-файла.



Обрабатывать все страницы

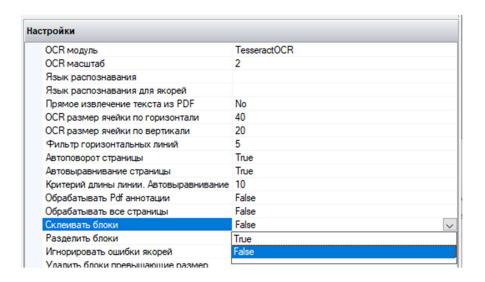
Обрабатывать все страницы – параметр, позволяющий включить обработку всех страниц в документе. Этот режим не делает проверку якоря LastPage.





Склеивать блоки

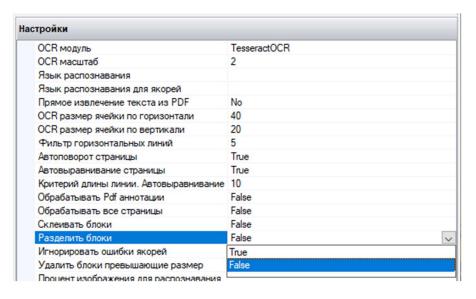
Склеивать блоки – параметр, позволяющий склеивать рядом стоящие блоки в один блок.



Разделить блоки

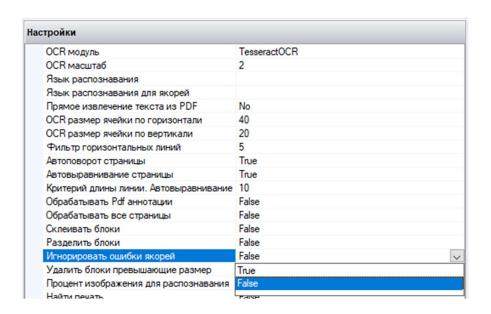
Разделить блоки – параметр, позволяющий разделить блоки содержащие пробелы на блоки без пробелов.





Игнорировать ошибки якорей

Игнорировать ошибки якорей – параметр, позволяющий отключить генерацию ошибок, если якори (любые) не найдены. В этом случае, область якоря считается нулевой.



Удалить блоки превышающие размер

Удалить блоки превышающие размер – параметр, позволяющий удалить блоки превышающие указанный размер. Поле для ввода расположено рядом с наименованием параметра. Если указать одно число, то будут удалены блоки, у которых ширина или высота будут больше этого значения. Также, можно указать значения через запятую в формате: ширина, высота.



OCR модуль	TesseractOCR
OCR масштаб	2
Язык распознавания	
Язык распознавания для якорей	
Прямое извлечение текста из PDF	No
OCR размер ячейки по горизонтали	40
OCR размер ячейки по вертикали	20
Фильтр горизонтальных линий	5
Автоповорот страницы	True
Автовыравнивание страницы	True
Критерий длины линии. Автовыравнивание	10
Обрабатывать Pdf аннотации	False
Обрабатывать все страницы	False
Склеивать блоки	False
Разделить блоки	False
Игнорировать ошибки якорей	False

Процент изображения для распознавания

Процент изображения для распознавания – параметр, позволяющий указать процент изображения, который пойдет на распознавание ОСR движком. Поле для ввода расположено рядом с наименованием параметра. Значение необходимо записывать как одно число или два числа через тире.

Например:

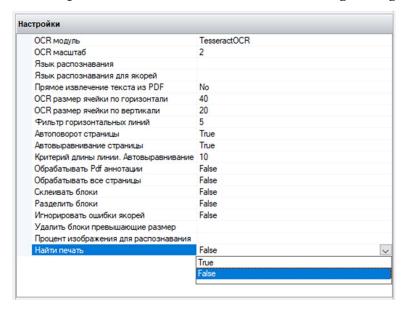
30 (0-30) - будет распознано 0-30% изображения; 30-70 - будет распознана часть изображения; 70-100 - будет распознано нижние 30% изображения.

OCR модуль	TesseractOCR
OCR масштаб	2
Язык распознавания	
Язык распознавания для якорей	
Прямое извлечение текста из PDF	No
OCR размер ячейки по горизонтали	40
OCR размер ячейки по вертикали	20
Фильтр горизонтальных линий	5
Автоповорот страницы	True
Автовыравнивание страницы	True
Критерий длины линии. Автовыравнивание	10
Обрабатывать Pdf аннотации	False
Обрабатывать все страницы	False
Склеивать блоки	False
Разделить блоки	False
Игнорировать ошибки якорей	False

Найти печать



Найти печать – параметр, позволяющий включить поиске печатей на документе. В атрибуты будет добавлен ключ "Stamps", который возвратит массив объектов StampItem со свойствами X, Y, Width, Height, PageIndex.



1.3.2 Якоря

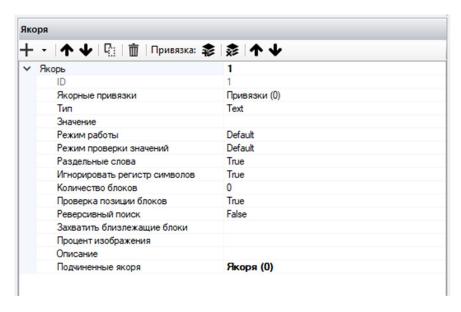
Якоря — блоки на обр аботанном документе, с помощью которых робот определяет, в какой части документа размещаются определенные данные.

При помощи якорей задаются границы определенных атрибутов. Якорей может быть неограниченное количество, в зависимости от типа документа.

Инструменты для работы с якорями

При создании нового якоря на рабочей панели открывается список свойств и инструментов для работы с якорем.





Основные инструменты для работы с якорями.

№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	кнопка + •	Позволяет добавить новый якорь. Кнопка * , которая находится рядом позволяет добавить новый подчиненный якорь, при этом нужно сделать клик на свойство "Подчиненные якоря" у того якоря, которому нужно добавить подчиненный якорь
2.	кнопка 🔨	Позволяет переместить якорь вверх.
3.	кнопка◆	Позволяет переместить якорь вниз.
4.	кнопка	Позволяет копировать выделенный якорь и автоматически добавить новый якорь вместе с заданными настройками.
5.	кнопка	Позволяет удалить выделенный якорь.
6.	кнопка 🕏	Позволяет добавить новую привязку к выделенному якорю.



7.	кнопка 🎉	Позволяет удалить выделенную привязку.
8.	кнопка 🔨	Позволяет переместить привязку вверх.
9.	кнопка♥	Позволяет переместить привязку вниз.



Настройки якоря

Параметры настроек якоря

№ п/п	Параметр	Значение параметра	
1.	ID	Идентификатор якоря. Присваивается автоматически.	
2.	Якорные привязки	Список якорных привязок.	
3.	Тип	 Тип якоря. Выбирается из выпадающего списка: Тext - текст; Table - таблица или ячейка таблицы; Packet - пакет взаимозаменяемых якорей (при привязке к этому якорю будут взяты его подчиненные якоря, первый успешно найденный из списка). 	



4	0	о п п
4.	Значение	Значение якоря. Для якоря типа Text возможно указывать подстановочные знаки (*)(?)
		указывать подстановочные знаки ()(:)
		Для якоря типа Table значение записывается в
		формате «Индекс таблицы:Индекс строки:Индекс
		колонки». Например, для Text «*обл* страх*» (два
		слова при включении «Раздельные слова»), для
		Table – «0:1:2» (таблица 0, строка 1, колонка 2),
		«1:2» (таблица 1, строка 2), «2» (вся таблица 2)
5.	Режим работы	Режим работы якоря. Возможные значения:
		• Default – якорь обязателен на первой
		странице многостраничного документа;
		• LastPage – якорь будет обязателен в
		многостраничном документе и укажет на
		последнюю страницу;
		• Optional - необязательный якорь и если он
		не будет найден на странице, то ошибка не
		генерируется.
6.	Режим	Выбор режима проверки значений. Возможные
	проверки	значения:
	значения	
		• Default (по умолчанию);
		• EveryWordContains (проверка каждого слова
		в значении как вхождение в строку.
		Пример: "тест" => "*тест*");
		• RemoveNonWordSumbols (при проверке
		удаляются все не числовые и буквенные
		символы).



7.	Раздельные слова	При включенной настройке значение якоря разбивается на слова и каждое слово ищется отдельно.
8.	Игнорировать регистр символов	При включенной настройке будет игнорироваться регистр символов.
9.	Количество блоков	Если включена настройка «Раздельные слова», то при поиске якоря возможна ситуация, когда в заданной области могут быть найдены несколько значений блоков, соответствующие заданным словам. Данное свойство позволяет ограничить максимальное количество слов между блоками.
10.	Проверка позиции блоков	При включенной настройке проверяется позиция блоков относительно координаты Y и берутся блоки с наименьшим значением.
11.	Реверсивный поиск	При включенной настройке поиск якоря будет идти снизу страницы.
12.	Захватить близлежащие блоки	Захват ближних блоков якоря. Позволяет захватить в якорь ближние блоки от уже захваченного значения. Например, если указать +2-1, то к уже захваченному якорю будут присоединены ближние блоки: 2 блока справа и 1 блок слева. Данное свойство может использоваться, когда не удается установить полное устойчивое значение для якоря, но часть его имеет устойчивое значение.



	T	
13.	Процент	Процент изображения, в котором будет
	изображения	производиться поиск якоря. Значение необходимо
		записывать как одно число или два числа через
		тире. Например:
		30 (0-30) - будут использоваться 0-30%
		изображения;
		30-70 - будут использоваться часть изображения;
		70-100 - будут использоваться нижние 30%
		изображения.
14.	Описание	Пользовательское описание якоря.
15.	Подчиненные	Якоря, напрямую зависящие от родительского
	якоря	якоря.
	якоря	
	якоря	якоря.
	якоря	якоря. Поиск подчиненных якорей будет происходит
	якоря	якоря. Поиск подчиненных якорей будет происходит только после того, как будет найден
	якоря	якоря. Поиск подчиненных якорей будет происходит только после того, как будет найден главный/родительский якорь. В случае, если в
	якоря	якоря. Поиск подчиненных якорей будет происходит только после того, как будет найден главный/родительский якорь. В случае, если в настройках родительского якоря указан режим
	якоря	якоря. Поиск подчиненных якорей будет происходит только после того, как будет найден главный/родительский якорь. В случае, если в настройках родительского якоря указан режим «Optional» (т.е. он не является обязательным), и
	якоря	якоря. Поиск подчиненных якорей будет происходит только после того, как будет найден главный/родительский якорь. В случае, если в настройках родительского якоря указан режим «Optional» (т.е. он не является обязательным), и если родительский якорь не был найден, то и
	якоря	якоря. Поиск подчиненных якорей будет происходит только после того, как будет найден главный/родительский якорь. В случае, если в настройках родительского якоря указан режим «Optional» (т.е. он не является обязательным), и если родительский якорь не был найден, то и поиск подчиненных якорей не будет осуществлен,
	якоря	якоря. Поиск подчиненных якорей будет происходит только после того, как будет найден главный/родительский якорь. В случае, если в настройках родительского якоря указан режим «Optional» (т.е. он не является обязательным), и если родительский якорь не был найден, то и поиск подчиненных якорей не будет осуществлен, они будут пропускаться. Эта параметр

1.3.3 Атрибуты

Атрибуты позволяют определить и обозначить области документа, данные из которых будут распознаваться роботом и импортироваться в таблицу «Результаты».

Основные инструменты для работы с атрибутами.



№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	кнопка +	Позволяет добавить новый атрибут.
2.	кнопка 🄨	Позволяет переместить атрибут вверх.
3.	кнопка♥	Позволяет переместить атрибут вниз.
4.	кнопка 🖫	Позволяет копировать выделенный атрибут.
5.	кнопка 🗖	Позволяет удалить выделенный атрибут.
6.	кнопка 🕏	Позволяет добавить новую привязку к выделенному атрибуту.
7.	кнопка 🕏	Позволяет удалить выделенную привязку.
8.	кнопка 🔨	Позволяет переместить привязку вверх.
9.	кнопка♥	Позволяет переместить привязку вниз.



Свойства атрибутов

При создании нового атрибута на рабочей панели открывается список свойств и инструментов для работы с атрибутом.

Свойства и инструменты для работы с атрибутом.

№ п/п	Элемент интерфейса	Описание элемента интерфейса



1.	Имя	Имя атрибута, которое будет отражено в результатах
2.	Якорные привязки	Список якорных привязок атрибута.
3.	Ограничение линиями	При включенной настройке осуществляется поиск ограничения линиями. Атрибут может быть ограничен с любой стороны, относительно линий, распознанных ОСК модулем. Эти линии отображаются при нажатии кнопки «ОСК Линии». Возможные значения: • None – отключено; • UpDown – сверху и снизу; • Up – сверху; • Down — снизу.
4.	Корректировать угол	При включенной настройке осуществляется корректировка угла наклона области. Область корректируется, если она соприкасается с границами таблицы. При включенном значении «Автовыравнивание» корректировку угла использовать не нужно.
5.	Привязка области	Привязка к области осуществляется по 4-м параметрам: координата X, координата Y, ширина, высота области. Исходная точка области - вверху слева.



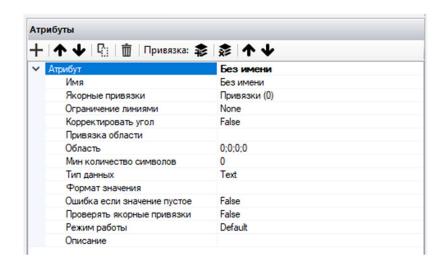
6.	Область	Прямоугольник области атрибута на странице.
7.	Мин количество символов	Минимальное количество символов, содержащихся в атрибуте. По умолчанию указано значение 0. Настройка используется, если атрибут должен содержать в себе определенное количество символов, например, атрибут ИНН. Если вычисленное значение меньше указанного будет сгенерирована ошибка.
8.	Тип данных	 Тип данных для атрибута. Возможные значения: Техt - текст; Int - число; Float - вещественное число; Date Time - дата и время; Chars - только буквы (исключаются цифры и другие символы); CountColorHSV - количество точек в заданном интервале цвета. Значение должно быть указано в формате «H1-H2;S1-S2;V1-V2», где H1, S1, V1 - это минимальные значения, а H2, S2, V2 - максимальные значения; Regex - регулярное выражение.



9.	Формат значения	Формат значения атрибута. Задается
		форматирование значения. Например:
		• если тип значения дата и время то,
		чтобы получить только дату,
		необходимо указать «dd.MM.yyyy»;
		• если тип значения CountColorHSV,
		то необходимо указать интервал
		цвета в формате «H1-H2;S1-S2;V1-
		V2». Формат указывается без
		кавычек;
		• Если тип значения Regex, то
		необходимо указать регулярное
		выражение.
10.	Ошибка если значение	Если при обработке шаблона значение
	пустое	атрибута пустое, то будет сгенерирована
		ошибка распознавания.
11.	Проверять якорные	Проверка якорных привязок атрибута.
	привязки	При включенной настройке будет
		производиться проверка якорных
		привязок. Если хоть одна из привязок не
		будет найдена, то атрибут, также, не будет
		найден.
12.	Режим работы	Режим работы атрибута. Возможные
	-	значения:
		• Default - стандартная работа
		атрибута (его значение будет
		заполнено при первом успешном
		определении якорей);



		 Аррend - добавление значения в атрибут при каждом успешном определении якорей на следующих страницах; Overwrite - перезапись, а не добавление значения атрибута.
13.	Описание	Пользовательское описание атрибута.



1.3.4 Таблицы

На рабочей панели справа отображаются все таблицы шаблона, которые были распознаны ранее. Их может быть неограниченное количество. Каждой таблице автоматически присваивается номер, нумерация начинается с 0. Также, отображается количество колонок каждой распознанной таблицы. Именно по этому значению можно определить с какой (или с какими) таблицами продолжить работу.

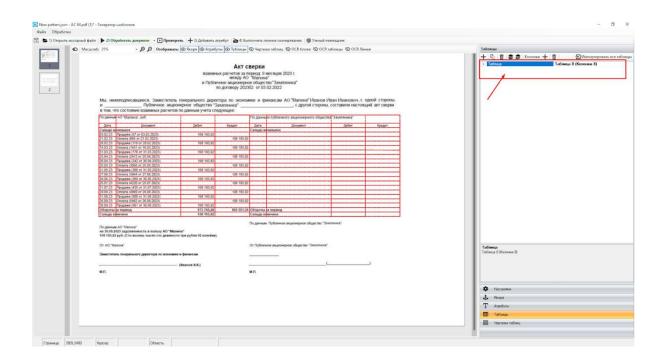
Основные инструменты для работы с таблицами.

№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	кнопка +	Позволяет добавить новую таблицу.



2.	кнопка С	Позволяет копировать выделенную таблицу.
3.	кнопка 🗖	Позволяет удалить выделенную таблицу.
4.	кнопка 🕏	Позволяет привязать якорь к выбранному свойству.
5.	кнопка 🏂	Позволяет удалить якорь из выбранного свойства.
6.	кнопка +	Позволяет добавить колонку таблицы.
7.	кнопка 🛅	Позволяет удалить колонку таблицы.
8.	кнопка Đ	Импортировать все таблицы из документа.







Свойства таблиц

На рабочей панели открывается список свойств и инструментов для работы с таблицей.

Свойства и инструменты для работы с таблицами.

№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	Индекс	Индекс OCR таблицы. Для улучшения
		привязки рекомендуется указывать
		минимальное количество колонок. В
		случае, если Шаблонизатор извлекает
		данные лишь из одной таблицы
		сканированного документа, необходимо
		указывать индекс, равный 0.
2.	Пропустить строки	При необходимости возможно указать
		количество первых строк, которые нужно
		пропустить для того, чтобы данные этих
		строк не попали в результаты работы
		робота. При выборе значения, равного 1,
		— первая строка будет пропущена.
		Наименования колонок будут
		соответствовать значениям первой строки.
		Если задать названия колонок в шаблоне,
		то эти значения не будут изменены.
3.	Строк в заголовке	При построении таблицы заданное
		количество строк будет использовано для
		создания заголовков колонок. При
		указании значения 0, названия будут взяты
		из шаблона колонок, в противном случае
		название будет суммироваться из ячеек



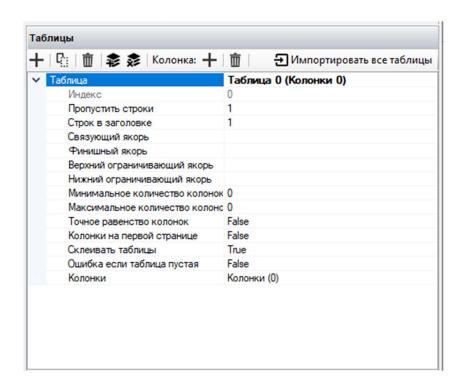
		колонки по заданному количеству строк. Например, если в заголовок должна попасть информация, размещенная в двух верхних строках таблицы, — необходимо указать количество строк в заголовке, равное 2.
4.	Связующий якорь	При работе с документом, в котором есть несколько таблиц можно указать связующий якорь, имеющий тип Optional. Якорь укажет на какие-либо слово, заголовок, пункт, после которых начинается нужная таблица. При нахождении такого якоря таблица, следующая за ним, будет распознаваться.
5.	Финишный якорь	Финишный якорь ограничивает извлечение таблицы до якоря, который может иметь тип Optional. Таким якорем, например, может служить строка в таблице "Всего к оплате". Таблица будет извлечена из документа до финишного якоря.
6.	Верхний ограничивающий якорь	Якорь, после которого таблица начнет строиться.
7.	Нижний ограничивающий якорь	Якорь, до которого таблица будет строиться.
8.	Минимальное количество колонок	Минимальное количество колонок в ОСК таблице. Все таблицы с количеством



		колонок меньше указанного пропускаются.
9.	Максимальное количество колонок	Максимальное количество колонок в ОСК таблице. Все таблицы с количеством больше указанного пропускаются.
10.	Точное равенство колонок	При включенной настройке извлекаться будут только те таблицы, у которых количество колонок в документе меньше или равно заданному в шаблоне.
11.	Колонки на первой странице	При включенной настройке колонки будут браться только с первой страницы. Данная настройка указывает на то, что на следующих страницах у таблицы нет заголовков.
12.	Склеивать таблицы	При включенной настройке таблицы с одинаковым количеством колонок будут склеиваться. Если настройка не включена, такие таблицы будут добавляться как новые.
13.	Ошибка если таблица пустая	При обработке шаблона, если таблица пустая, то будет сгенерирована ошибка распознавания.
14.	Колонки	Колонки таблицы. По каждой таблице отображается соответствующее ей количество колонок. Также, здесь будут указываться: • Индекс (задается автоматически);



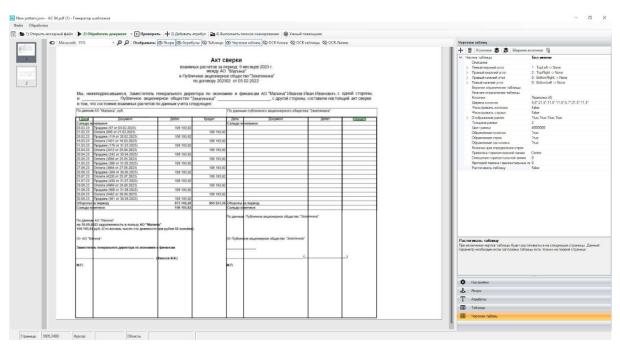
• Имя (задается пользователем);
• Тип данных (Text – текст; Int –
число; Float - вещественное число; Date
Time – дата и время; Chars – только буквы
(исключаются цифры и другие символы)).
В случае, если в результатах полного
сканирования отсутствует несколько
колонок, то мы можем их добавить.



1.3.5 Чертежи таблиц

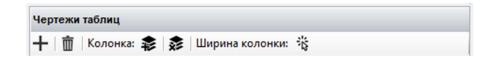
Чертежи таблиц применяются для документов, в которых по каким-либо причинам отсутствуют границы таблиц. Это могут быть иностранные документы (например, Invoice), в которых линии таблиц не предусмотрены. Также это могут быть российские документы, в которых частично не прорисованы линии таблицы, как на скриншоте ниже (отсутствуют межстрочные линии).





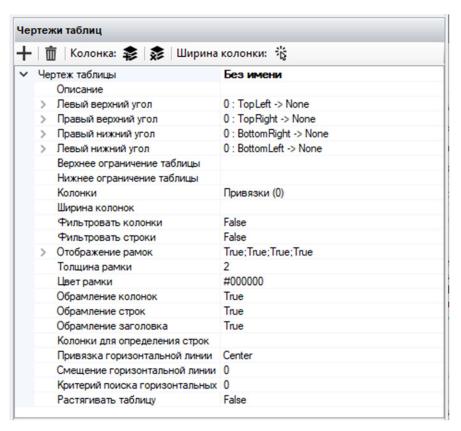
Основные инструменты для работы с чертежами таблиц.

№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	кнопка +	Позволяет добавить чертеж таблицы.
3.	кнопка 🗖	Позволяет удалить чертеж таблицы.
4.	кнопка 🕏	Позволяет добавить привязку колонки.
5.	кнопка 🕏	Позволяет удалить привязку колонки.
6.	кнопка	Позволяет добавить ширину колонки.



На рабочей панели справа отображаются все параметры, которые позволяют произвести настройку чертежа таблицы.





Свойства и инструменты для работы с чертежами таблиц.

№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	Описание	Описание чертежа таблицы.
2.	Левый верхний угол	Якорная привязка левого верхнего угла.
3.	Правый верхний угол	Якорная привязка правого верхнего угла.
4.	Правый нижний угол	Якорная привязка правого нижнего угла.
5.	Левый нижний угол	Якорная привязка левого нижнего угла.
6.	Верхнее ограничение таблицы	Ограничение таблицы сверху страницы. Можно указать значения в пикселях или процент, используя символ *. Например, 200*.
7.	Нижнее ограничение таблицы	Ограничение таблицы снизу страницы. Можно указать значения в пикселях или



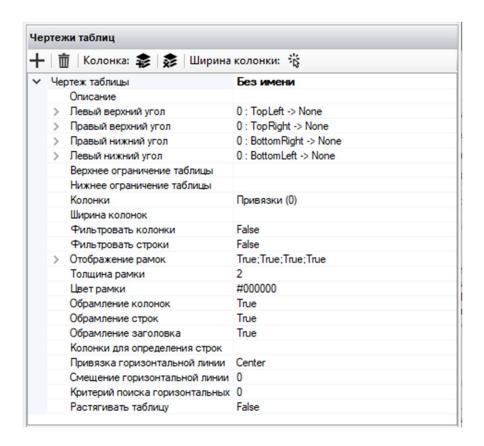
		процент, используя символ *. Например, 200*.
8.	Колонки	Якорные привязки для колонок.
9.	Ширина колонок	Ширина колонок. Можно указать фиксированное значение или процент от всей ширины таблицы в формате " 50*", " 10*". Вся ширина таблицы равна 100. Если сумма значений процентов всех указанных колонок равна или больше 100, то ширина таблицы делится пропорционально указанным значениям. Иначе, последняя колонка будет заполняться автоматически до значения 100.
10.	Фильтровать колонки	При включенной настройке для распознавания таблицы будут учитываться только те линии, которые нарисованы в чертежи таблицы.
11.	Фильтровать строки	При включенной настройке для распознавания таблицы будут учитываться только те линии, которые нарисованы в чертеже таблицы.
12.	Отображение рамок	Управление отображением рамок чертежа таблицы.
13.	Толщина рамки	Толщина рамки таблицы в пикселях.
14.	Цвет рамки	Цвет рамки таблицы в формате RGB. Например, «#000000» — черный цвет, «#FF0000» — красный цвет.



15.	Обрамление колонок	При включенной настройке отображаются линии колонок на чертеже таблицы.
16.	Обрамление строк	При включенной настройке отображаются линии строк на чертеже таблицы.
17.	Обрамление заголовка	При включенной настройке отображается первая строка на чертеже страницы.
18.	Колонки для определения строк	Индексы колонок, по которым будут определяться строки. Если значение будет пустое, то будут участвовать все колонки. Индекс начинается с 0.
19.	Привязка горизонтальной линии	Параметр привязки горизонтальной линии. Возможные значения: • Сепter - по центру между блоками текста; • Тор - верх нижнего блока текста; • Воttom - низ верхнего блока текста.
20.	Смещение горизонтальной линии	Смещение горизонтальной линии по оси Y в пикселях. Для точного смещения следует указывать значения в формате + Например, +10-20.
21.	Критерий поиска горизонтальных линий	Смещение блоков текста для определения их пересечения. По умолчанию значение равно 10.
22.	Растягивать таблицу	При включенной настройке чертеж таблицы будет растягиваться на следующие страницы. Данный параметр



необходим если заголовки таблицы есть только на первой странице.



Для осуществления привязки границ таблицы для каждого угла чертежа необходимо указать якорь и задать параметры расположения линий относительно выбранного якоря.



Параметры якорных привязок углов таблицы.

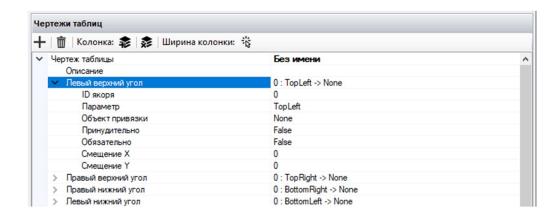
№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	ID якоря	ID якоря, к которому происходит привязка.
2.	Параметр	 Параметр привязки. Возможные значения: ТорLeft – привязка левого верхнего угла якоря к выбранному объекту привязки; ТорRight – привязка правого верхнего угла якоря к выбранному объекту привязки; ВоttomLeft – привязка левого нижнего угла якоря к выбранному объекту привязки; ВоttomRight – привязка правого нижнего угла якоря к выбранному объекту привязки; Left – привязка левой грани якоря (посередине) к выбранному объекту привязки Тор – привязка верхней грани якоря (посередине) к выбранному объекту привязки; Right – привязка правой грани якоря (посередине) к выбранному объекту привязки; Воttom – привязка верхней грани якоря (посередине) к выбранному объекту привязки; Воttom – привязка верхней грани якоря (посередине) к выбранному объекту привязки;



		 LeftArea - привязка всей левой области якоря к выбранному объекту привязки; RightArea - привязка всей правой области якоря к выбранному объекту привязки; ВottomArea - привязка всей нижней области якоря к выбранному объекту привязки; ТорArea — привязка всей верхней области якоря к выбранному объекту привязки; FullArea - привязка якоря со всех сторон к выбранному объекту привязки.
3.	Объект привязки	Точка или грань области атрибута для привязки выбранного параметра.
4.	Принудительно	При включенной настройке указанный параметр якоря будет задан к выбранному объекту привязки принудительно и без проверок.
5.	Обязательно	При включенной настройке ошибка данной привязки поиск всего элемента будет считаться неудачным.
6.	Смещение Х	Смещение по оси X в пикселях. Для точного смещения следует указывать значения в формате + Например, +10-20. Данная настройка используется в случае, когда после якоря указывается какой-либо



		символ (скобка или двоеточие), который не должен попасть в область атрибута.
7.	Смещение Ү	Смещение по оси Y в пикселях. Для точного смещения следует указывать значения в формате + Например, +10-20.



Для отображения рамок чертежа таблицы необходимо задать параметры отображения каждой из рамок.

Параметры отображения рамок чертежа таблицы.

№ п/п	Элемент интерфейса	Описание элемента интерфейса
1.	Левая рамка	Отображение левой рамки чертежа таблицы.
2.	Верхняя рамка	Отображение верхней рамки чертежа таблицы.
3.	Правая рамка	Отображение правой рамки чертежа таблицы.
4.	Нижняя рамка	Отображение нижней рамки чертежа таблицы.



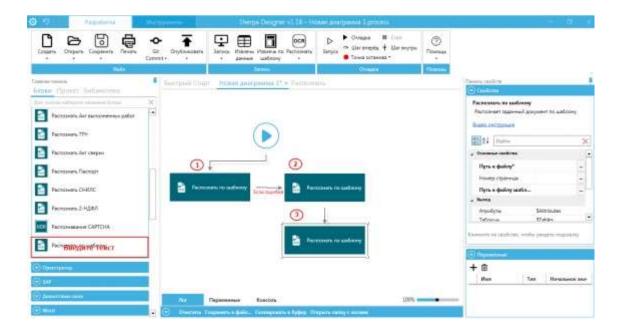
 Отображение рамок 	True;True;True;True
Левая рамка	True
Верхняя рамка	True
Правая рамка	True
Нижняя рамка	True

2. Действия в Sherpa Designer после создания шаблона

2.1 Значения переменных

После создания шаблонов для каждого типа документа (их может быть несколько 5, 10 и т.д.), можно пойти двумя путями:

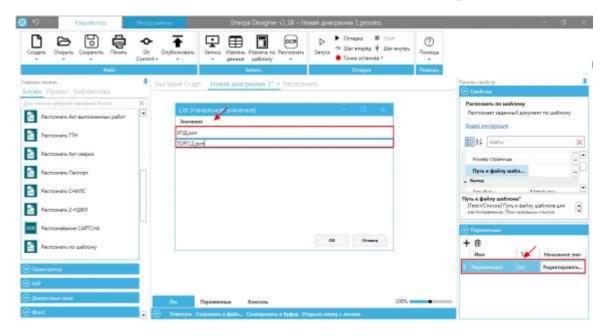
1) Настроить работу сценария проекта с поочередным распознаванием по каждому шаблону, при этом добавлять в сценарий столько блоков, сколько было создано шаблонов



2) Запустить распознавание по списку шаблонов в одном блоке.

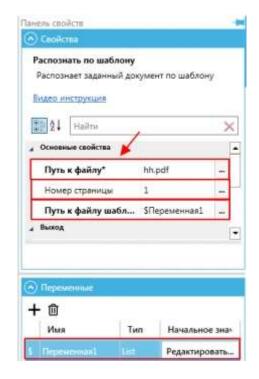
Прежде чем использовать список, необходимо его задать в Переменных. Выбираем Тип переменной List (список), задаем имя переменной и редактируем значение переменной: добавляем в список названия созданных нами шаблонов, по которым будет производиться распознавание. Например, ТОРГ12.json, УПД.json и т.д.





На панели свойств указываем путь к файлу PDF. Файл может быть многостраничным. Также поддерживаются файлы формата .png и .jpg.

Далее указываем путь к файлу шаблона для распознавания. При указании списка файлов шаблонов будут последовательно использоваться каждый файл, пока распознавание не закончится успешно. Если ни один из шаблонов не подойдет для документа, то будет сгенерировано исключение с ошибкой первого шаблона из списка.





Указываем путь к файлу в формате .pdf

В строке Путь к файлу шаблона указанную созданную нами переменную со списком шаблонов.

Номер страницы – 1, тогда распознавание начнется с первой страницы документа.

При запуске роботы блока «Распознать по шаблону», первым распознается документ, указанный в списке Переменной. В случае возникновения ошибки по одному из шаблонов, робот переходит к распознаванию следующего шаблона.

При использовании списка возможно увеличение скорости обработки нескольких шаблонов, если параметры распознавания страницы у этих шаблонов одинаковые.

Если параметры распознавания различны, то время распознавания документа увеличивается.

Например, если для одного из шаблонов указан ОСК-модуль Tesseract, а для другого шаблона – Yandex Vision (или разный масштаб или язык), то для каждого из этих шаблонов робот будет заново распознавать документ, т.к. указаны разные параметры.

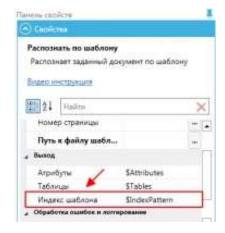
Поэтому рекомендуем при создании шаблонов для типовых стандартизированных документов оставлять настройки, заданные по умолчанию и менять их только в случае необходимости.

В случае, если параметры настроек некоторых шаблонов были изменены, то рекомендуем компоновать их по группам: сначала указать несколько шаблонов с одним видом настроек, затем несколько шаблонов с другим типом настроек.

Индекс шаблона

После распознавания документа будет определен индекс шаблона.

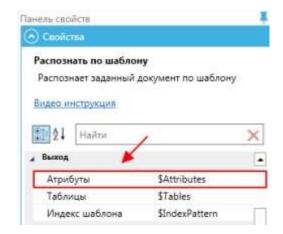




Используя блок «Условие» мы задаем траекторию дальнейшего сценария согласно полученным индексам. (например, если индекс 0 – то это УПД, если индекс 1 – Счет-фактура и т.д.)

Если для распознавания использовался один документ, а не список, то индекс по умолчанию всегда будет 0 (нулевой).

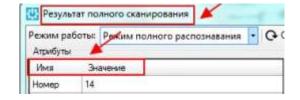
Атрибуты



После обработки шаблона необходимо продолжить работу с созданными атрибутами.

Атрибуты отображаются в окне «Результаты полного сканирования» Шаблонизатора (*Имя, Значение*).





Чтобы получить значения атрибутов мы должны обратиться к словарю и присвоить значение переменной. Используем блок «Присвоить значение переменной».

На панели свойств выбираем значение переменной, кликаем на «...» и редактируем выражение. В открывшемся окне выбираем категорию \$Attributes.

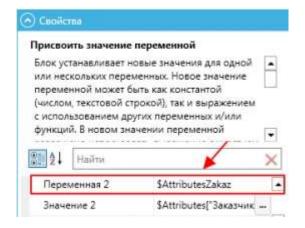


Указываем Имя атрибута, которое ранее было задано в Шаблонизаторе, используя кавычки.



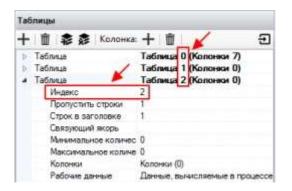
Далее необходимо задать имя переменной, с которой мы продолжим работать. Имя переменной может быть задано произвольно, например, \$AttributesZakaz или \$MyOrganization.





Таблица

Присваивая значение переменной для таблицы (если таблиц несколько, — то массив таблиц), необходимо использовать индекс таблицы, который был присвоен в Шаблонизаторе автоматически (по порядку). Вручную присвоить значение переменной невозможно.



На панели свойств выбираем значение переменной, кликаем на «...» и редактируем выражение. В открывшемся окне выбираем категорию \$Tables. Добавляем столько таблиц, сколько было добавлено в Шаблонизаторе и указываем индекс каждой таблицы.





Далее необходимо задать имя переменной, с которой мы продолжим работать. Имя переменной может быть задано произвольно, например, \$TablesProducts или \$MyTables.



После того, как Имя и Значение переменных были заданы, можно приступать к дальнейшей работе.

2.2 Работа с многостраничными документами

При создании шаблонов для многостраничных документов необходимо определить уникальные якоря, поиск по которым будет осуществляться только в определенном типе документа. Например, в документе УПД может встречаться слово счет-фактура, соответственно, не нужно использовать это слово в качестве якоря ни для УПД, ни для Счет-фактуры.

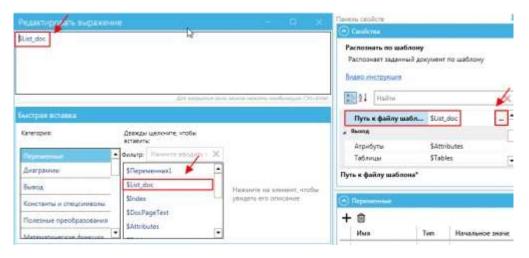
При необходимости работы с многостраничным документом, бывает, что один документ находится на нескольких страницах.

Для каждого типа документов необходимо создавать отдельный шаблон (например, 1 — Торг12, 2 – Счет-фактура, 3 – УПД), и указать все типы документов в значениях переменной. При этом тип переменной выбрать List (список).

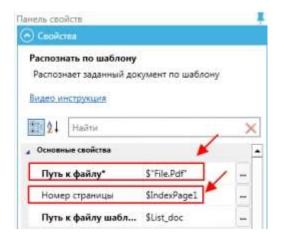


Далее указываем путь к файлу шаблона:





Указываем Номер страницы – 1, для того чтобы распознавание начиналось с первой страницы. Также указываем путь к файлу.



Завершение работы робота при возникновении ошибки

По окончании работы по создания шаблона, необходимо прописать дальнейший сценарий его распознавания и вывода результатов.

Если робот не смог произвести распознавание документа, то сценарий может быть остановлен. Для этого в Sherpa Designer добавляем блок «Конец». Также можно добавить блок «Лог» для записи ошибки в лог.



Ошибка распознавания может возникать в случаях, когда ни один из шаблонов не подошел или, когда не осталось документов для распознавания. 64



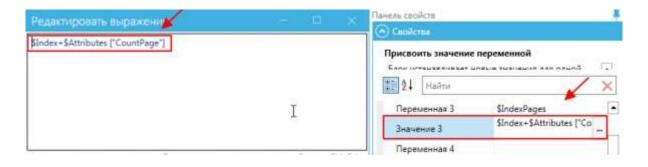
В случае, если в PDF файле находятся несколько документов, и один из типов документов неизвестен роботу (т.е. на этот тип документа нет шаблона), робот будет пропускать этот документ и будет распознавать лишь те документы, для которых есть шаблоны.

Поиск второго и последующих документов в многостраничном файле

После того, как в многостраничном документе был распознан один из шаблонов, необходимо внести изменения в прописанный индекс (прибавить), чтобы дальнейшее распознавание началось не со следующей страницы, а после найденного документа.

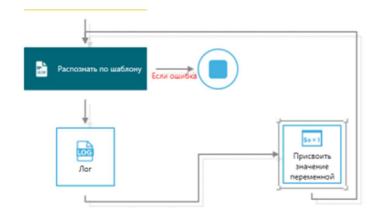
Так как в атрибутах существует параметр CountPage, который соответствует количеству страниц, участвующих в распознавании шаблоном, то для продолжения процесса распознавания нужно добавить этот параметр к индексу.

Таким образом, если в многостраничном документе был распознан первый документ, занимающий несколько страниц, при добавлении параметра CountPage дальнейшее распознавание начнется со страницы, следующей после распознанного документа.



После этого настраиваем сценарий на повторное распознавание документа.

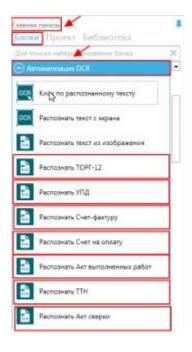




3. Примеры

3.1 Создание шаблона на примере документа «Справка о стоимости выполненных работ»

В случае, когда одним из действий сценария является распознавание и импорт данных из сканированного документа, необходимо воспользоваться блоками из вкладки Автоматизация ОСR, которые находятся на главной панели Дизайнера (Sherpa Designer).



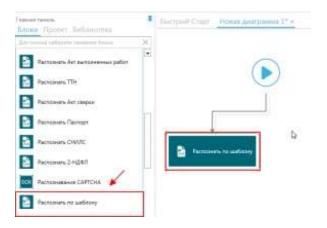
В Дизайнере есть предустановленные шаблоны по следующим видам документов:

- Торг12;
- УПД;
- Счет-фактура;



- Счет на оплату;
- Акт выполненных работ;
- TTH;
- Акт сверки;
- Паспорт;
- СНИЛС;
- НДФЛ;
- CAPCHA.

В случае, когда роботу необходимо распознать информацию из иного вида документа, не указанного во вкладке «Автоматизация ОСК», потребуется создать шаблон документа. При этом в сценарий проекта добавляется блок «Распознать по шаблону».



По созданному шаблону в дальнейшем робот будет распознавать и обрабатывать информацию из всех документов данного вида (Справка о стоимости выполненных работ и затрат, Акт об оказании услуг и т.д.).

Для каждого вида документов создается отдельный шаблон.

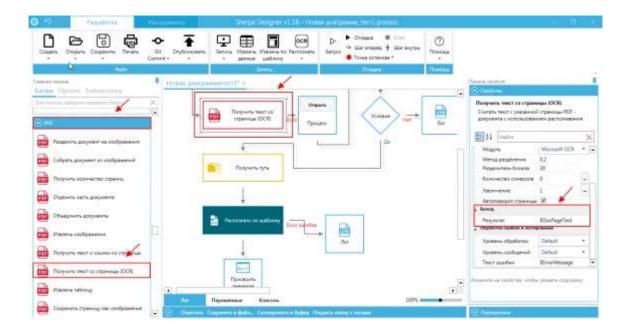
3.1.1 Предобработка документа

Рассмотрим создание шаблона на примере документа «Справка о стоимости выполненных работ и затрат».

Прежде чем робот приступит к созданию шаблона, документ должен быть распознан роботом и сохранен по указанному пути.

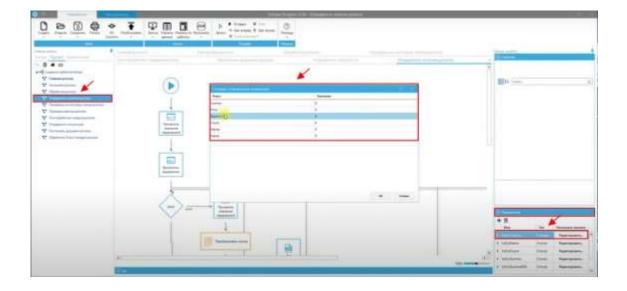


Например, при распознавании документа в формате .pdf в проекте сценария добавляем блок «Получить текст со страницы ОСК». В настройках результатов во вкладке «Выход» указываем \$DocPageText



Важно понимать, что в некоторых распознаваемых файлах название столбцов таблиц может не совпадать с названием колонок, которые заданы для вывода результата роботом. Также некоторые столбцы могут отсутствовать вовсе или могут отсутствовать некоторые границы таблиц.

Проверить и сопоставить названия столбцов необходимо в сценарии проекта «Определить колонки.process», они задаются в настройках справа: «Панель свойств» — «Переменные».





Например, после обработки документа «Справки о стоимости работ и затрат» в результате роботом должны выводиться следующие колонки:

Summa	сумма
Price	цена
SummaNDS	сумма с НДС
Stavka	ставка
Name	имя/название
Count	количество

Но при создании шаблона мы видим, что в самом документе часть этих данных отсутствует.

3.1.2 Создание шаблона

Создание шаблона происходит на верхней панели основного меню Дизайнера во вкладке «Разработка» — раздел «Запись» — «Извлечь по шаблону».



Извлечь по шаблону

При нажатии на кнопку «Извлечь по шаблону» открывается обработка «Генератор шаблонов» (Шаблонизатор).

При создании нового шаблона происходит автоматическое распознавание документа. Так как мы создаем шаблон документа, с которым ранее не работали, нам необходимо создать новые якоря и привязки к тем значениям, которые будут обрабатываться роботом именно в данном типе документа. Поэтому на сообщение во всплывающем окне «Вы хотите импортировать все таблицы из документа в шаблон» отвечаем «Нет».





Распознавание текста

Распознавание текста в Шаблонизаторе производится с помощью встроенных и внешних модулей распознавания:

- Tesseract OCR
- ABBYY Cloud OCR
- ABBYY Fine Reader
- Microsoft OCR
- YandexVision

В платформе Sherpa интегрированы несколько ОСR -модулей. Два из них - офлайн (поставляются вместе с роботом), и будут работать без подключения к интернету. Это Tesseract OCR и Microsoft OCR.

Yandex Vision и ABBYY OCR – онлайн-модули, т.е. используют функции соответствующих облачных модулей.

ABBYY Fine Reader – это коммерческий офлайн-модуль, использование которого требует отдельной лицензии.

Платформа позволяет настроить работу сценария с распознаванием изображений и переключаться между этими движками в любой момент времени. Для этого необходимо выбрать желаемый движок из списка.



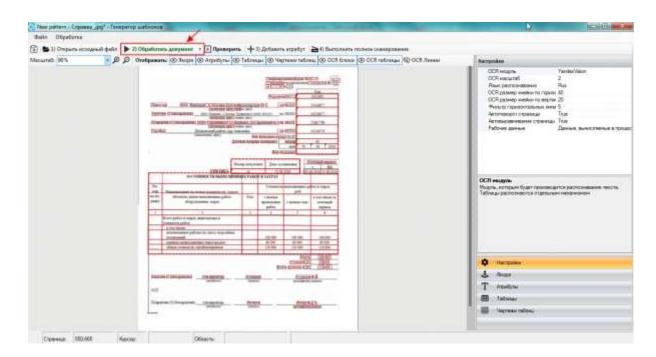




В случае, если распознаваемый документ высокого качества, с высоким расширением, то достаточно использовать бесплатные ОСR-модули.

При обработке документов низкого качества лучше использовать платные решения.

После того, как основные параметры настроек заданы, документ необходимо обработать. Действие запускается нажатием кнопки «Обработать документ».

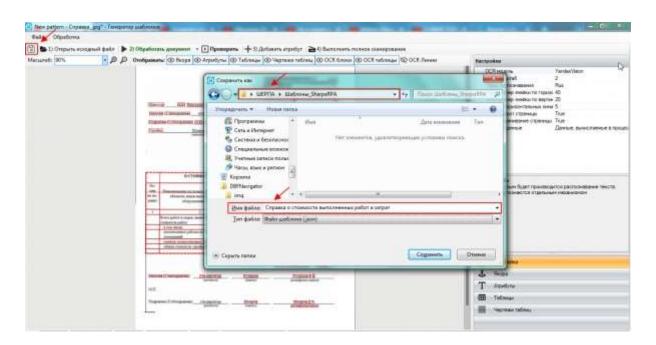


Обработанный документ выглядит так, как показано на скриншоте выше.

Сохранить шаблон



Для дальнейшей работы с шаблоном его нужно сохранить, нажав на иконку в левом верхнем углу основной панели действий. Далее указываем путь для сохранения шаблона и имя файла. Имя файла указываем исходя из вида документа, например, «Справка о стоимости выполненных работ и затрат».



3.1.3 Работа с таблицами

После сохранения шаблона переходим к работе с таблицами.

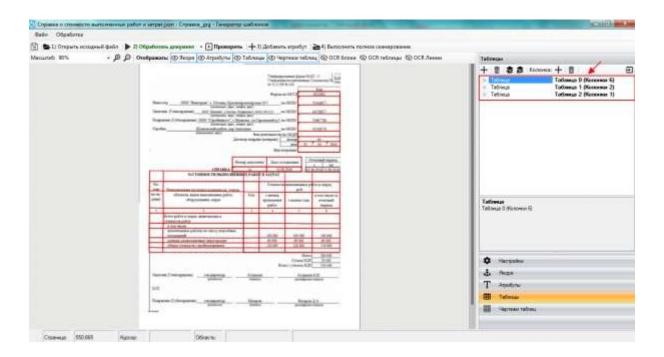




В левой части основной панели настроек выбираем вкладку «Таблицы» — нажимаем кнопку «Импортировать» 🗐.

Импортируем все таблицы документа, для этого в появившемся окне «Вы хотите импортировать все таблицы документа в шаблон?» выбираем «Да».

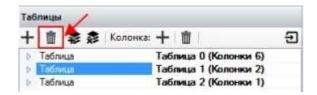
На рабочей панели справа отобразятся все таблицы шаблона, которые были распознаны. Их может быть одна-две и более. В нашем случае на скриншоте ниже мы видим, что ОСR-модулем распознаны 3 таблицы шаблона.



Каждой таблице автоматически присваивается номер, нумерация начинается с 0. Также отображается количество колонок каждой распознанной таблицы. Именно по этому значению мы можем определить с какой (или с какими) таблицами мы продолжим работать.

В нашем примере нас интересуют данные из таблицы с 6-ю колонками, соответственно, остальные таблицы (с 2 и с 1 колонками) мы можем удалить из окна панели.

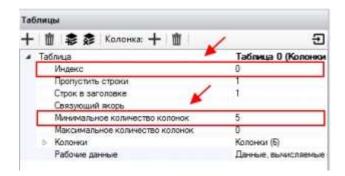




Удалить таблицу можно нажав на значок на верхней панели окна, предварительно выделив таблицу, которую нужно удалить.

Индекс

Индекс ОСR-таблицы. В случае, если Шаблонизатор извлекает данные лишь из одной таблицы сканированного документа, необходимо указывать индекс, равный 0. При этом для улучшения привязки рекомендуется указывать минимальное количество колонок.



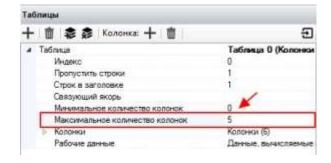
Минимальное количество колонок

Указывается минимальное количество колонок в ОСR-таблице. Все таблицы с количеством меньше указанного будут пропущены.

Максимальное количество колонок

Указывается максимальное количество колонок в ОСR-таблице. Все таблицы с количеством больше указанного будут пропущены.





При создании шаблона «Справки о стоимости выполненных работ и затрат» указываем минимальное количество колонок равное 5, т.к. нас интересуют данные из таблицы с 6 колонками. Остальные таблицы данного документа содержат менее 5 колонок.

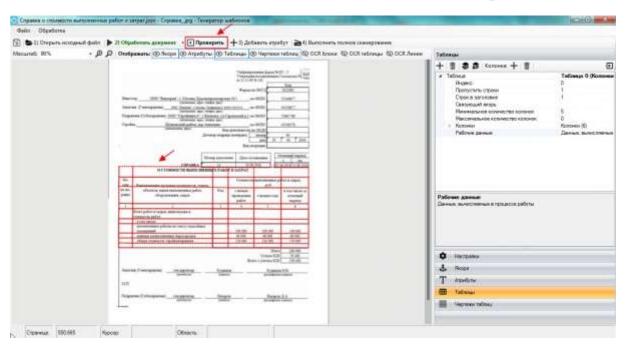
Таким образом Шаблонизатор продолжит работу с интересующей нас таблицей.

При создании шаблонов для других типов документов (с большим количеством таблиц и колонок в таблицах) можно указывать и максимальное и минимальное количество колонок, тем самым задавая интервал, в который будут попадать таблицы с интересующими нас данными.

Проверить

Для проверки корректности отображения табличной части, из которой будут импортироваться данные нажимаем кнопку «Проверить» на верхней рабочей панели.





Также проверку можно запустить нажатием клавиши F5.

Пропустить строки

Указываем количество строк, которые нужно пропустить для того, чтобы данные этих строк не попали в результаты работы робота.

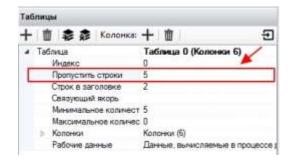
При выборе значения, равного 1, — первые строки будут пропущены. Наименования колонок будут соответствовать значениям первой строки.

Если задать названия колонок в шаблоне, то эти значения не будут изменены.

В нашем примере необходимо пропустить 5 верхних строк таблицы, т.к. значения интересующих нас данных указаны в таблице документа начиная с 6 строки.

Для этого в строке «Пропустить строки» указываем число 5.

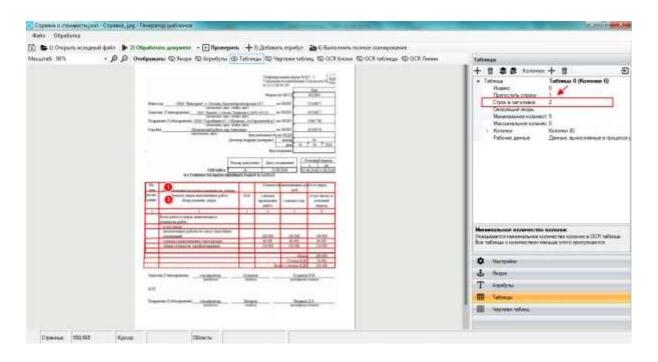




Строк в заголовке

При построении таблицы, заданное количество строк будет использовано для создания заголовков колонок. При указании значения 0 заголовки будут взяты из шаблона колонок, в противном случае название будет суммироваться из ячеек колонки по заданному количеству строк.

Например, в шаблоне «Справки о стоимости выполненных работ и затрат» необходимо указать количество строк заголовке равное 2, т.к. в заголовок должна попасть информация, размещенная в двух верхних строках таблицы.

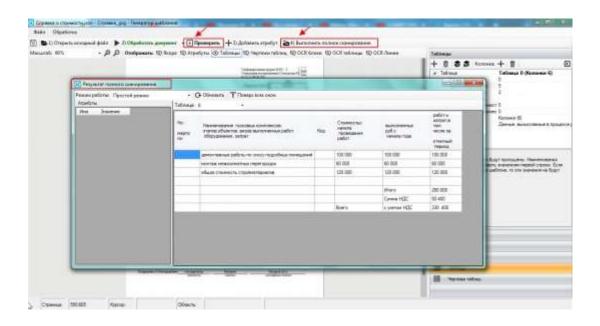


После выполненных действий необходимо проверить промежуточный результат. Нажимаем «Проверить» — далее «Выполнить полное сканирование».

Результат полного сканирования



Результат полного сканирования отобразится в виде таблицы с импортированными данными из табличной части сканированного документа.

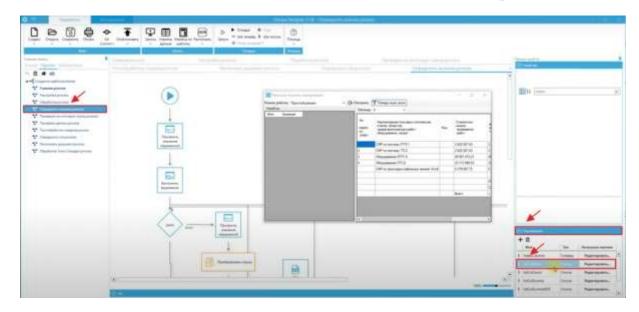


Определить колонки

На этом этапе часть данных может отражаться некорректно, в частности, названия колонок могут не совпадать с теми значениями, которые ранее были заданы в сценарии проекта «Определить колонки.process» — в настройках «Панель свойств» — «Переменные».

Для корректировки выводимых результатов необходимо вернуться в сценарий проекта «Определить колонки.process», сопоставить значения переменных, заданных в настройках «Панель свойств», добавить указанные здесь колонки и определить для них названия.

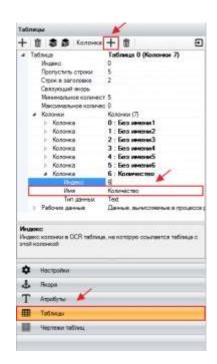




Добавление колонок «Количество», «НДС Сумма», «Ставка».

В нашем примере в результатах полного сканирования отсутствует несколько колонок, в том числе колонка «Количество».

Название колонки должно соответствовать значению, присвоенному данной переменной «количество».







Добавление недостающих колонок производится на правой рабочей панели Шаблонизатора во вкладке «Таблицы». Нажимаем на кнопку +, добавляем колонку, присваиваем колонке имя «Количество».

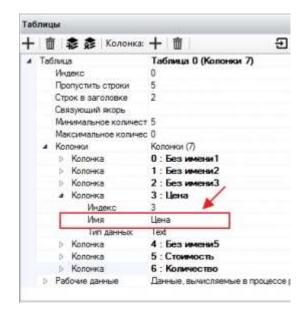
При повторной проверке в результатах полного сканирования Шаблонизатора появится колонка «Количество».

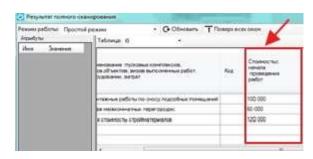
Производим проверку остальных колонок и корректируем их названия в зависимости от значений переменных в сценарии проекта «Определить колонки.process».

При создании шаблона «Справка о стоимости выполненных работ и затрат» аналогичным образом добавляем колонки: «НДС Сумма», «Ставка».

В случае, если название колонки не соответствует названию переменной, но содержание колонки соответствует значению переменной, то создавать дополнительную колонку не нужно, достаточно изменить название колонки в Шаблонизаторе.





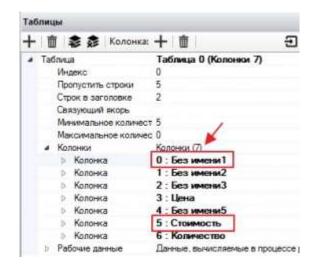


Например, в таблице с результатами Шаблонизатора название колонки «Стоимость начала проведения работ», а название переменной «Цена», тогда мы вносим изменения в название колонки, предварительно убедившись в том, что именно в этой колонке отображается цена.

Аналогичным образом необходимо изменить названия всех колонок, которые не совпадают с названиями переменных.

Нумерация колонок таблицы в Шаблонизаторе начинается с 0.





В случае если название колонки сохранено по умолчанию и содержит в себе числовое значение (например, «Без имени1»), то в таблице результатов Шаблонизатора название указанной колонки определится автоматически и будет содержать в себе текст из соответствующей колонки сканированного документа.

Если название колонки мы изменили вручную, то в таблице результатов Шаблонизатора будет отображаться именно заданное нами название колонки.

Поэтому, рекомендуем обратить особое внимание на часть табличного документа, в котором название колонки указано в объединенных ячейках. Даже если при создании шаблона документа название колонки из объединенных ячеек было распознано корректно, в других документах название может быть указано со смещением.

В нашем примере слово «Стоимость» располагается в 4 ячейке таблицы, но в некоторых документах может сместиться в 5 ячейку. В таких случаях рекомендуем переименовать все ячейки, названия которых могут быть смещены.





По завершению работы с таблицей в Шаблонизаторе необходимо произвести проверку и удостовериться в корректности результатов полного сканирования.

3.1.4 Якоря

Точки на сканированном документе, с помощью которых робот определяет, в какой части документа размещаются определенные данные.

При помощи якорей мы задаем границы определенных атрибутов. Якорей может быть сколь угодно много, в зависимости от типа документа.



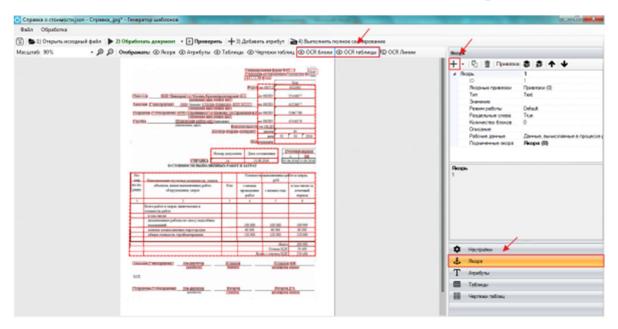
В нашем примере необходимо получить данные следующих атрибутов:

- «Инвестор»
- «Заказчик (Генподрядчик)»
- «Номер документа»
- «Дата»

Создание якоря (атрибут «Инвестор»)

Рассмотрим процесс создания якоря на примере атрибута «Инвестор».





Для создания якоря переходим во вкладку «Якоря» в правой рабочей области Шаблонизатора и нажимаем кнопку в верхней части панели. В этот момент становятся активны «ОСК блоки» и «ОСК таблицы» для удобства выбора определенного блока в качестве якоря.

Для того, чтобы Шаблонизатор мог корректно определить местоположение атрибута в документе, необходимо ограничить его границы. Ограничение верхней границы задается по тексту, находящемуся на строку выше самого атрибута. В нашем случае это блок «Форма по».

Для присвоения значения якорю достаточно кликнуть левой клавишей мыши по ОСК-блоку (слову), которое мы выбрали в качестве якоря. Также можно указать значение вручную, написав якорное слово в строке «Значение».

В случае, если мы укажем в качестве якоря слово «Форма», то якорь будет установлен неверно, т.к. в документе есть несколько ОСR-блоков, распознавших это слово.

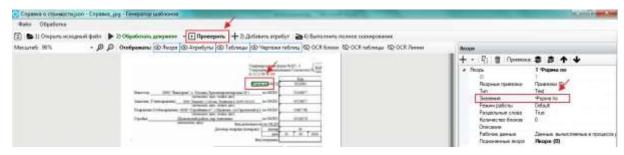


Поэтому меняем указанное значение якоря на «Форма по». После этого необходимо произвести проверку, чтобы убедиться, что якорное слово определилось корректно.



Когда значение якоря задается нажатием клавишей мыши на ОСR-блок, — к значению добавляется знак «*», если значение указывается вручную, — добавлять этот знак необязательно. При использовании якорных привязок знак «*» нужно удалить и после этого добавить привязку.

В случае, если значение содержит два и более слова, — между ними необходимо ставить пробел, т.к. он выступает разделителем слов.



Далее добавляем якорь для атрибута «Инвестор». Для корректного распознавания информации об Инвесторе, необходимо ограничить якорными точками данный атрибут слева и справа.

Слева якорем будет выступать слово «Инвестор». Добавляем якорь, нажимаем левой клавишей мыши на ОСК-блок «Инвестор» и Шаблонизатор автоматически заполняет строку «Значение», тем самым определяя якорную точку.



Справа якорем будет выступать слово «по». По аналогии добавляем якорь в рабочей области Шаблонизатора, нажимаем на ОСК-блок «по» и производим проверку.

После проверки видим, что якорь установлен неверно, т.к. в документе есть несколько ОСR-блоков, распознавших слово «по».

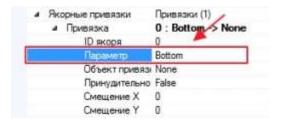
Для корректного определения якоря осуществляем якорную привязку к якорю «Форма по», который находится строкой выше того блока, который мы хотим указать в качестве якоря. Для этого нажимаем на кнопку «Добавить привязку», затем кликаем по якорю «Форма по», тем самым привязываем наш



новый якорь к ОСR-блоку, который располагается строкой выше нужного нам якоря.



Для того, чтобы Шаблонизатор определил координаты создаваемого якоря относительно якоря «Форма по» указываем параметр привязки «Bottom». Этот параметр говорит о том, что создаваемый якорь будет привязан к якорю «Форма по» и находится строкой ниже.

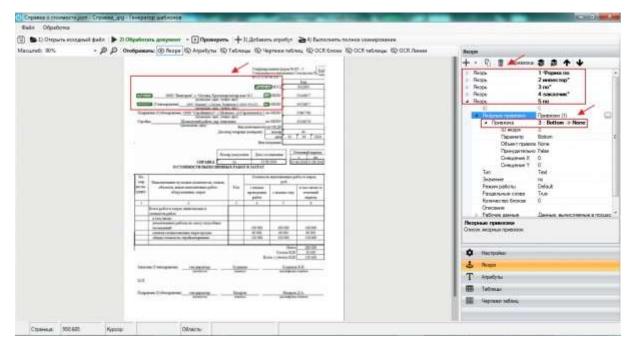


Создание якоря (атрибут «Заказчик»)

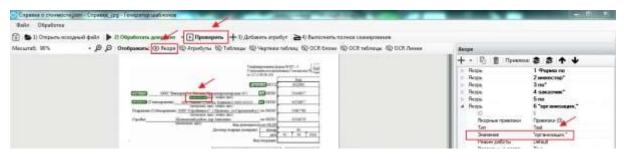
По аналогии создаем якоря для атрибута «Заказчик». Слева якорем будет выступать слово «Заказчик». Добавляем якорь, нажимаем левой клавишей мыши на ОСR-блок «Заказчик» и Шаблонизатор автоматически заполняет строку «Значение», определяя якорную точку.

Справа якорем будет выступать слово «по». По аналогии добавляем якорь в рабочей области Шаблонизатора, нажимаем на ОСR-блок «по», производим проверку и добавляем якорную привязку. В этом случае привязку осуществляем к якорному слову «по», которое располагается строкой выше в документе. Указываем параметр привязки «Bottom» и производим проверку.



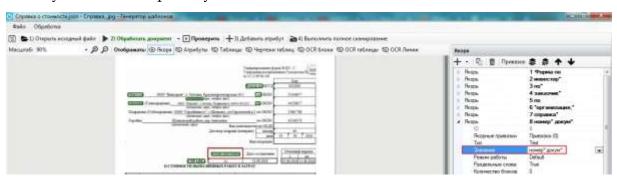


Атрибут «Заказчик» необходимо ограничить не только справа и слева, но и сверху. Для этого в качестве якоря выбираем слово «организация». Создаем еще один якорь, редактируем название так, чтобы с обеих сторон слова стоял знак «*» — «*организация*». Проверяем корректность расположения якорей.



При создании якорей, в значении которых содержится несколько слов, необходимо указывать знак «*» в конце каждого слова.

Если сочетание слов встречается в документе единожды, то второе слово можно написать сокращенно. Например, в качестве якоря по номеру документа можем указать «номер* докум*».



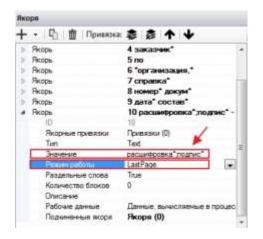


Если в начале значения якоря присутствует лишний символ (чаще всего это скобка), мы можем удалить этот символ и вместо него поставить знак «*». Тогда получим значение *организация* вместо (организация*.

Завершающий якорь

При создании шаблона необходимо указать якорь, который будет завершать документ. Т.е. по этому якорю мы можем определить, что в документе таблиц больше нет и работу с таблицами можно завершать.

В нашем примере завершающим якорем будет слово расшифровка.



В нашем примере завершающим якорем будет слово расшифровка*.

Шаблонизатор позволяет задать для якоря несколько альтернативных вариантов, по которым робот будет распознавать завершающий атрибут.

Добавим слово **подпис***, укажем намеренно без окончания, т.к. в документе может быть указано как подпись, там и подписи.

Если в документе отсутствует слово расшифровка*, то робот будет искать слово nodnuc*

Для завершающего якоря указываем *Режим работы – LastPage*.

3.1.5 Создание атрибутов

Следующий этап в создании шаблона - создание атрибутов.

Атрибуты — позволяют определить и обозначить области документа, данные из которых будут распознаваться роботом и импортироваться в таблицу «Результаты».

Атрибуты «Инвестор» и «Заказчик»





В нашем примере в таблицу с результатами должны импортироваться данные об Инвесторе, Заказчике, а также Дата и Номер документа

В Шаблонизаторе необходимо добавить эти атрибуты на рабочей панели справа: выбираем «Атрибуты», нажимаем кнопку добавить, присваиваем имя атрибута «Инвестор».

Для корректного распознавания области документа, данные из которой будут импортироваться в результаты, для атрибута необходимо осуществить привязку к ранее созданным якорям.



В нашем примере якорем справа будет выступать якорь «Инвестор», слева - якорь «по».

Якорные привязки к атрибуту «Инвестор»

Для создания привязок в созданном атрибуте нажимаем «Якорные привязки», далее «Добавить привязку», затем левой клавишей мыши кликаем на



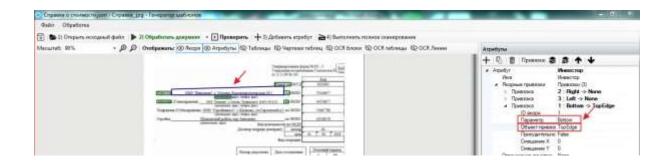
якорное слово «Инвестор». Аналогичным образом создаем привязку к якорному слову «по». Производим проверку, в результате которой область данных нашего атрибута должны подсветиться синей рамкой, как на скриншоте выше.

Чтобы ограничить область данных атрибута сверху, создаем третью якорную привязку к ОСR-блоку «Форма по». Нажимаем «Добавить привязку», кликаем левой клавишей мыши на созданный ранее якорь «Форма по». В параметрах привязки выбираем значение «Bottom», т.к. в качестве ориентира мы будем использовать нижнюю границу данного якоря.



Помимо параметра в привязке для данного атрибута необходимо указать «Объект привязки». Т.к. данные атрибута могут располагаться выше строки, которую мы выделили при помощи первых двух якорей слева и справа.

В строке «Объект привязки» указываем «ТорЕdge» — верхняя грань.



Якорные привязки к атрибуту «Заказчик»

По аналогии создаем атрибут «Заказчик», устанавливаем привязки к ранее созданным якорям «Заказчик» справа, «по» слева и «организация» сверху. Для привязки к якорю сверху устанавливаем параметр «Bottom», «Объект привязки» указываем «ТорЕdge».



В результате по каждому атрибуту должно быть указано по три якорные привязки: справа, слева и сверху.

После проверки в шаблоне каждый атрибут будет отображаться в виде синей рамки. В случае, если границы рамки не захватывают всю область данных атрибута, необходимо внести изменения в настройки.



Ограничение линиями

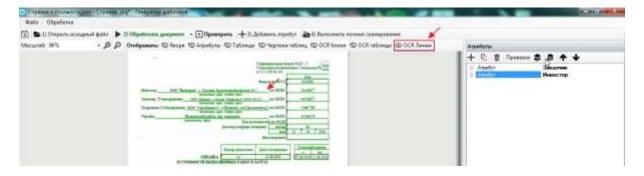
В некоторых документах названия полей, которые мы используем в качестве якорей могут располагаться не на нижней линии с описанием поля, а сверху или посередине, как на рисунке ниже.



В таких случаях для корректной привязки рекомендуем воспользоваться дополнительным параметром «Ограничение линиями». Предварительно нужно проверить присутствуют ли в шаблоне ОСR-линии, ограничивающие созданные нами атрибуты.

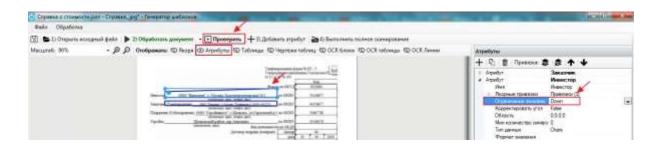
Нажимаем кнопку OCR-линии на верхней рабочей панели и проверяем наличие линий.





Линии подсвечены зеленым цветом, и мы можем использовать их в качестве дополнительного ориентира в работе с атрибутами.

На верхней рабочей панели выбирает «Атрибуты», кликаем левой клавишей мыши на один из атрибутов, после чего он становится активным и в правой рабочей панели открываются инструменты для работы с выбранным атрибутом.

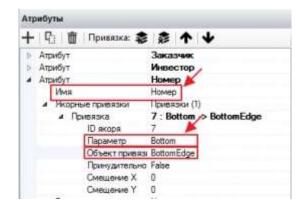


Выбираем «Ограничение линиями», из выпадающего списка выбираем значение «Down». Таким обратом мы ограничиваем нижнюю границу атрибута, которая теперь будет ориентироваться на ОСК-линии.

Обязательно проверяем корректность отображения границ атрибута. При необходимости меняем значения данного параметра.

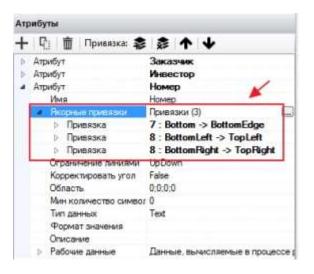
Создание атрибутов «Номер» и «Дата». Якорные привязки





Добавляем в правой рабочей панели атрибут, задаем имя атрибута *Номер*. Добавляем привязки к ранее созданным якорям.

В качестве привязки слева используем якорь *Справка**, в качестве параметра выбираем *Воttom*, также указываем *Объект привязки – BottomEdge*.



Создаем вторую привязку к якорю *номер* докум**. В качестве параметра выбираем *BottomLeft* (нижний левый угол). Объект привязки указываем *TopLeft* (верхний левый угол).

Таким же образом производим привязку атрибута к правому углу. Создаем третью привязку также к якорю *номер* докум**. В качестве параметра выбираем *BottomRight* (нижний правый угол). Объект привязки указываем *TopRight* (верхний правый угол).

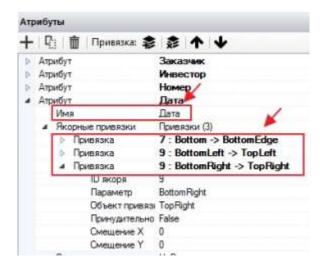
Таким образом, для атрибута Номер должны быть созданы три якорные привязки.



После проверки в шаблоне должен появиться синий прямоугольник, внутри которого отображены данные нашего атрибута.



Создание атрибута Дата.



Добавляем в правой рабочей панели атрибут, задаем имя атрибута *Дата*. Добавляем привязки к ранее созданным якорям.

В качестве привязки слева используем якорь Справка*, в качестве параметра выбираем *Bottom*, также указываем *Объект привязки – BottomEdge*.

Создаем вторую привязку к якорю дата* *состав**. В качестве параметра выбираем *BottomLeft* (нижний левый угол). Объект привязки указываем *TopLeft* (верхний левый угол).

Таким же образом производим привязку атрибута к правому углу. Создаем третью привязку также к якорю дата* $cocma\theta$ *. В качестве параметра выбираем BottomRight (нижний правый угол). Объект привязки указываем TopRight (верхний правый угол).

В результате для атрибута *Номер* должны быть созданы три якорные привязки.

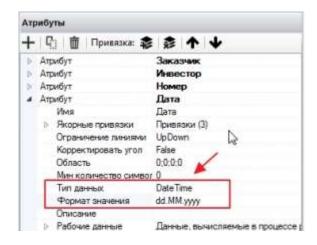


После проверки в шаблоне должен появиться синий прямоугольник, внутри которого отображены данные нашего атрибута.



Тип данных - DateTime.

Формат значения – вручную указываем формат, необходимый для атрибута. В нашем примере это *dd.MM.yyyy*



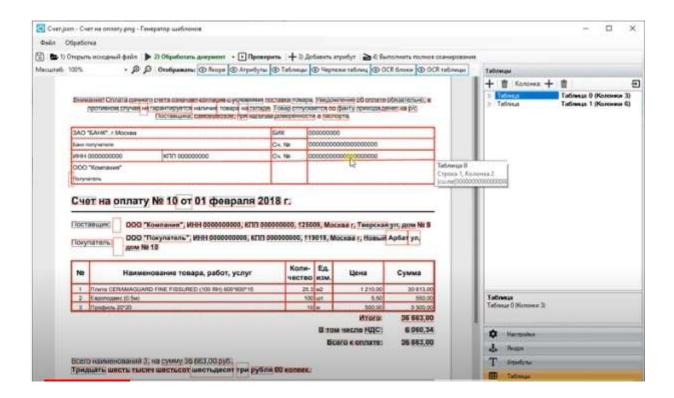
В случае, если в каком-либо документе дата документа будет указана как 30 ноября 2022г., в результате эта дата будет преобразована в 30.11.2022.

3.2 Создание шаблона с фиксированными атрибутами на примере документа «Счет на оплату»

Фиксированные атрибуты имеют положение и размеры, привязанные только к сторонам документа (масштабу), но не привязанные к конкретным якорям, поэтому их можно свободно перемещать по документу. Фиксированные атрибуты допустимо использовать только если во всех возможных образцах документа необходимые для извлечения данные находятся примерно на одном и том же месте и не смещаются по документу.



При создании шаблона «Счет на оплату», Шаблонизатор распознает все таблицы, которые содержатся в документе. Эта информация отображается на вкладке «Таблицы».

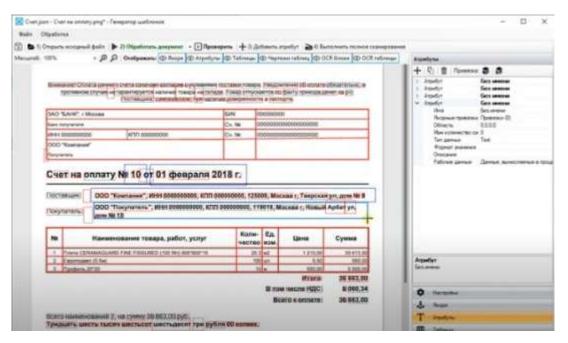


В случае, если роботом будут использоваться данные лишь из одной таблицы, вторая таблица может быть удалена. Для этого достаточно выделить ненужную таблицу и нажать на кнопку « в верхней части рабочей области «Генератора шаблонов».

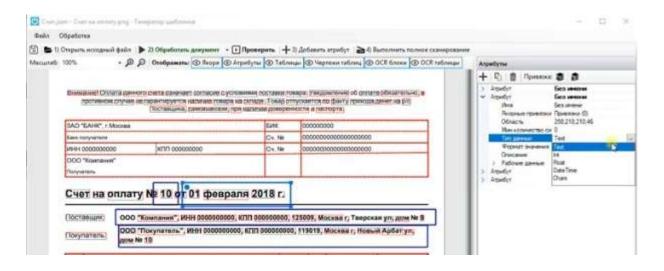


Вкладка «Атрибуты» позволяет определить и обозначить области документа, данные из которых будут распознаваться роботом и импортироваться в документы «Результаты».





Для каждого атрибута (выделенной области, данные из которой будут импортированы) необходимо определить Тип данных из выпадающего списка. Например, для атрибута «дата документа» выбираем «Date Time».



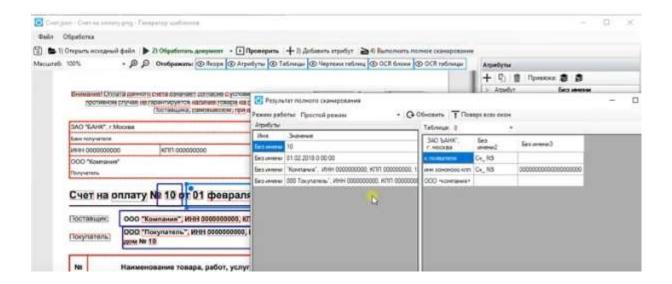
При необходимости любой атрибут может быть удален или добавлен новый.

После завершения работы во вкладке «Атрибуты» производится проверка шаблона и выполняется полное сканирование. Нажимаем «Проверить», затем «Выполнить полное сканирование».

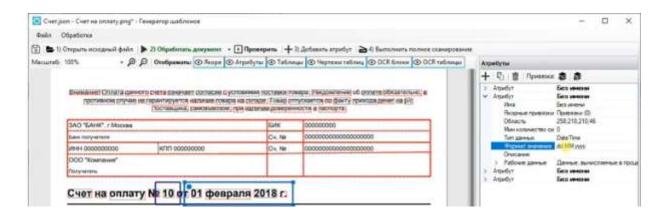
Открывается окно «Результат полного сканирования», в котором отображаются распознанные данные из документа в виде заполненной таблицы.



Если при проверке обнаружены неточности, то их в любой момент можно исправить.



При выборе типа данных «дата документа» дополнительно можно указать «Формат значения» вручную. После чего провести повторную проверку и убедиться в корректности значений.



При распознавании данных двух и более таблиц обрабатываемого документа, в окне «Результат полного сканирования» будут отображаться значения всех таблиц. Для просмотра значений по всем таблицам предусмотрена возможность переключения таблиц в выпадающем списке.

