

ABBYY FlexiCapture 9.0

Руководство по настройке проектов

Уважаемый пользователь!

Данное руководство содержит описание действий, которые необходимо выполнить при настройке проекта.

Для более детального ознакомления вы можете обратиться к другой справочной литературе:

- Контекстная справка. Вызывается из меню программы или с помощью клавиши F1. Также доступна из меню Старт > Программы > ABBYY FlexiCapture 9.0 Stations > Справки.
- Руководство для системного администратора ABBYY FlexiCapture 9.0. Старт > Программы > ABBYY FlexiCapture 9.0 Stations > Руководства > Руководство для системного администратора.
- Руководство для операторов. Старт > Программы > ABBYY FlexiCapture 9.0 Stations > Руководства > Руководство оператора.
- Руководство по созданию машиночитаемых форм. Старт > Программы > ABBYY FlexiCapture 9.0 Stations > Руководства > Создание машиночитаемых форм.
- Контекстная справка по приложениям FlexiLayout Studio 9.0 и FormDesigner 9.0 вызывается из меню соответствующего приложения или с помощью клавиши F1. Также доступна из меню Старт > Программы > ABBYY FlexiCapture 9.0 Stations > Справки.
- Руководство ABBYY FlexiLayout Studio «Обучение на примерах». Старт > Программы > ABBYY FlexiCapture 9.0 Stations > Руководства > Обучение на примерах.

Для автономной версии в путях вместо ABBYY FlexiCapture 9.0 Stations следует читать ABBYY FlexiCapture 9.0.

Желаем приятной работы с нашим продуктом!

Содержание

| | | |
|------------|--|----|
| 1. | Введение | 4 |
| 1.1. | Задача ввода данных и документов | 4 |
| 1.2. | Автоматизация ввода данных и документов | 4 |
| 1.3. | Какие документы можно обрабатывать с помощью ABBYY FlexiCapture 9.0..... | 5 |
| 1.4. | Станции ABBYY FlexiCapture 9.0 | 6 |
| 2. | Создание проекта | 6 |
| 3. | Создание определений документов..... | 7 |
| 3.1. | Создание жесткой разметки | 8 |
| 3.1.1. | Геометрическая разметка объектов определения..... | 8 |
| 3.1.1.1. | Группы полей | 9 |
| 3.1.1.2. | Поля без разметки | 10 |
| 3.1.1.3. | Разметка таблиц | 10 |
| 3.1.1.4. | Поля с несколькими областями | 11 |
| 3.1.1.5. | Поля с несколькими экземплярами | 11 |
| 3.1.1.6. | Исключение региона из распознавания | 12 |
| 3.1.1.7. | Удаление полей | 12 |
| 3.1.2. | Статические элементы..... | 12 |
| 3.1.2.1. | Особенности штрих-кода..... | 13 |
| 3.1.3. | Свойства полей | 14 |
| 3.1.3.1. | Общие свойства полей | 14 |
| 3.1.3.2. | Типы данных | 15 |
| 3.1.3.2.1. | Типы данных текстового поля ввода..... | 15 |
| 3.1.3.2.2. | Типы данных для меток | 18 |
| 3.1.3.2.3. | Типы данных для групп меток..... | 19 |
| 3.1.3.3. | Свойства распознавания полей..... | 21 |
| 3.1.3.3.1. | Свойства распознавания текстового поля ввода..... | 21 |
| 3.1.3.3.2. | Свойства распознавания меток и групп меток..... | 22 |
| 3.1.3.3.3. | Свойства распознавания штрих-кода | 23 |
| 3.1.3.3.4. | Свойства распознавания картинки | 24 |
| 3.1.3.4. | Параметры верификации | 24 |
| 3.1.3.5. | Параметры экспорта картинок | 25 |
| 3.1.3.6. | Проверка с помощью правил..... | 25 |
| 3.1.3.7. | Пользовательский скрипт | 27 |
| 3.1.4. | Как создать определение для многостраничных документов..... | 28 |
| 3.1.5. | Как создать определение документа с приложениями..... | 30 |
| 3.1.6. | Настройка экспорта данных | 31 |
| 3.1.6.1. | Экспорт в файл..... | 32 |
| 3.1.6.2. | Экспорт в базу данных..... | 32 |
| 3.1.6.3. | Экспорт в библиотеку документов SharePoint | 33 |
| 3.1.6.4. | Экспорт изображений | 34 |
| 3.1.6.5. | Экспорт, описанный с помощью скрипта | 35 |
| 3.1.7. | Настройка отображения распознанных данных | 35 |
| 3.1.8. | Тестирование определения документа | 35 |
| 3.1.9. | Редактирование и публикация определения документа | 36 |
| 3.2. | Настройка системы для ввода гибких документов | 36 |
| 3.2.1. | Классификаторы | 37 |
| 3.3. | Особенности ввода неструктурированных документов | 38 |

| | | |
|--------|--|----|
| 4. | Настройка свойств проекта..... | 39 |
| 5. | Настройка типов пакетов..... | 39 |
| 5.1. | Общие свойства..... | 40 |
| 5.2. | Обработка изображений..... | 40 |
| 5.3. | Распознавание..... | 41 |
| 5.4. | Проверка целостности..... | 41 |
| 5.5. | Экспорт..... | 41 |
| 5.6. | Процесс обработки..... | 42 |
| 5.6.1. | Автономная версия..... | 42 |
| 5.6.2. | Распределенная версия..... | 42 |
| 6. | Настройка импорта изображений..... | 43 |
| 7. | Загрузка проекта на сервер..... | 44 |
| 8. | Горячие кнопки..... | 45 |
| 8.1. | Главное окно приложения..... | 45 |
| 8.2. | Окно редактора определений документов..... | 47 |
| 8.3. | Окно групповой верификации..... | 48 |
| 8.4. | Окно контекстной верификации..... | 49 |

1. Введение

1.1. Задача ввода данных и документов

Сегодня различные виды документов используются повсеместно в бизнесе, на производстве, в сфере услуг. Заявления, анкеты, счета, платежные поручения и множество других документов играют важную роль в деятельности любой организации. При современном уровне развития информационных технологий нельзя обойтись только лишь документами на бумаге, и большинство данных с бумажных носителей должны быть введены в электронную систему для хранения, анализа и дальнейшей обработки.

Основной сложностью при обработке бумажного документа является ввод данных в компьютерную систему. Традиционный, ручной способ ввода документов может быть оправдан при небольшом объеме данных. Однако при больших объемах данных ручной ввод становится неэффективным из-за огромных трудозатрат, организационных издержек и невысокой скорости. Производительность ручного ввода при необходимости не может быть оперативно увеличена, группа ручного ввода сложно масштабируется, а финансовые и временные затраты на ее расширение сопоставимы с затратами на организацию всего процесса «с нуля».

Таким образом, ручной ввод данных не является оптимальным. Альтернативный способ ввода данных – менее сложный и гораздо более эффективный – применение системы автоматизированного ввода данных и документов. Такой системой и является ABBYY FlexiCapture 9.0.

1.2. Автоматизация ввода данных и документов

ABBYY FlexiCapture 9.0 – программное решение для потокового ввода данных из документов различного типа – структурированных (жестких форм), частично структурированных (гибких документов) и неструктурированных документов.

Автоматизированный ввод данных осуществляется в несколько этапов:

- пачка страниц **сканируется** при помощи документного сканера,
- отсканированные страницы автоматически комплектуются в документы,
- символы автоматически **распознаются**,
- символы, относительно которых у системы распознавания сложилось несколько гипотез (то есть однозначность их распознавания вызывает некоторые сомнения), подаются на **проверку** оператору,
- окончательно подтвержденные данные **экспортируются** в файл или базу данных, а изображения документов сохраняются в указанной папке. Изображения можно сохранять в графических форматах, а также в PDF с возможностью поиска.

Использование системы ABBYY FlexiCapture 9.0 является эффективным решением задачи автоматизации ввода, позволяющее легко управлять процессами, контролировать качество вводимых данных.

1.3. *Какие документы можно обрабатывать с помощью ABBYY FlexiCapture 9.0*

ABBYY FlexiCapture 9.0 – программное решение для потокового ввода данных с документов различного типа. Документы разного типа могут обрабатываться в одном потоке, также можно настроить обработку документов смешанного типа.

Рассмотрим подробнее разные виды документов, которые можно обрабатывать с помощью ABBYY FlexiCapture 9.0.

- **Структурированные документы.** Структурированными называются документы с набором выделенных информационных полей, оформление, количество и местоположение которых не меняются от одного экземпляра документа к другому. Как правило, такие формы распространяются в виде бланков и заполняются от руки. Чтобы идентифицировать жесткую форму в потоке документов и извлечь из нее данные, необходимо создать жесткую разметку, которая укажет программе места расположения полей, из которых нужно извлечь данные. Создание жесткой разметки осуществляется непосредственно в программе ABBYY FlexiCapture 9.0 при настройке проектов.
- **Частично структурированные документы.** К ним относятся документы, содержащие набор информационных полей, оформление, количество и расположение которых могут значительно отличаться в разных экземплярах документа. Эти документы называются гибкими. К примеру, счета являются гибкими документами, т.к. они часто отличаются как количеством позиций, так и форматированием, потому что они поступают от разных компаний. На всех счетах присутствуют номер счета и сумма платежа, но и они расположены в разных местах. Для идентификации гибких документов и извлечения из них данных программа ABBYY FlexiCapture 9.0 использует гибкую разметку (FlexiLayout). Для создания гибкой разметки используется приложение ABBYY FlexiLayout Studio, подробную информацию о работе которого можно найти в руководстве и справке к этому модулю. Процесс обработки гибких документов отличается от обработки жестких форм только на этапе создания и подключения разметки.
- **Неструктурированные документы.** Если вам нужно обрабатывать неструктурированные документы, содержащие информацию, изложенную в свободном стиле, например, контракты, письма, приказы, графики, то и в этом случае поможет ABBYY FlexiCapture 9.0.

Неструктурированные документы могут быть автоматически идентифицированы программой либо как приложения к жестким или гибким документам, либо с помощью гибкой разметки, а затем экспортированы в виде PDF файлов с возможностью поиска или файлов графических форматов. Извлекать индексные поля из неструктурированных документов можно как автоматически с помощью гибкой разметки, так и с помощью ручного ввода. Типичным сценарием обработки неструктурированных документов является задача перевода бумажного архива в электронный вид с необходимостью извлечения двух-трех индексных полей для организации быстрого поиска по атрибутам.

1.4. Станции ABBYY FlexiCapture 9.0

В зависимости от типа установки (распределенная или автономная) в состав системы входят различные станции. При автономном типе установки настройка проекта осуществляется на станции администратора. При распределенной установке настройка проекта осуществляется на станции настройки проектов, после чего он выгружается на сервер приложений.

2. Создание проекта

Проект включает в себя все необходимые настройки для потокового ввода документов (определения документов, профили импорта изображений, настройки программы) и обрабатываемые документы.

Документы объединяются в **пакеты**. Способ формирования пакетов зависит от принципа обработки, например, можно объединять в пакеты документы, введенные одной и той же датой или из отсканированного за раз набора документов. Настройки, с которыми будут обрабатываться пакеты документов, определяются типом пакета. Можно создать несколько типов пакетов с наиболее часто используемыми наборами настроек обработки. Тогда, чтобы указать с какими настройками должен обрабатываться пакет, оператору необходимо просто задать тип пакета. Также можно указать настройки обработки для пакета в целом. Эти настройки будут действовать для пакетов с предустановленным типом «По умолчанию». При создании сначала указываются настройки проекта, а потом на их основе формируются типы пакетов (настройки проекта являются настройками «по умолчанию» для типов пакетов).

Для обработки документов используются **рабочие пакеты**. **Тестовые пакеты** используются в процессе отладки определений документов. Отличие этих пакетов в том, что для обработки документов из тестовых пакетов используется локальная (неопубликованная) версия определения, а для обработки документов из рабочих пакетов – опубликованная версия. К списку тестовых пакетов можно перейти непосредственно из окна редактора определения документа.

Документ состоит из изображения одной или нескольких страниц (одностраничные и многостраничные документы) и извлеченных из них данных.

В проект можно включать несколько определений документов, тогда документы разных типов будут обрабатываться в рамках одного проекта. Таким образом, при использовании программы, отпадает необходимость предварительно сортировать обрабатываемые документы, документы разных типов можно вводить в одном потоке. Тем не менее, если потоки документов не пересекаются, то для них можно создать отдельные проекты.

На первом этапе настройки нужно создать проект и хотя бы одно определение документа.

Чтобы создать проект, нажмите кнопку **Новый...** в диалоге **Открытие проекта**, который появляется при запуске программы, либо выберите пункт меню **Файл > Новый проект...** в основном окне программы. Выберите директорию для хранения проекта и введите его название.

Для добавления пакетов в проект используйте пункт контекстного меню **Новый пакет**, появляющегося при нажатии правой кнопки мыши в главном окне со списком пакетов. Вы можете загружать изображения, не создав ни одного пакета, в этом случае пакет будет создан программой автоматически.

Чтобы увидеть документы, добавленные в пакет, щелкните дважды левой кнопкой мыши на названии пакета. Вернуться к списку пакетов можно, выбрав пункт меню **Проект > Рабочие пакеты** или нажав комбинацию клавиш **Ctrl+B**.

3. Создание определений документов

Создание определения документа – основной этап настройки проекта, от которого зависит качество данных, получаемых в результате обработки. Для того чтобы создать определение документа, нужно выполнить следующие действия:

- создать разметку (жесткую – непосредственно в редакторе определения документа или загрузив форму, разработанную в ABBYY FormDesigner, гибкую – загрузив файл, разработанный в приложении ABBYY FlexiLayout Studio 9.0);
- задать свойства каждого поля: указать, какие именно данные нужно искать в каждом поле (это существенно улучшает качество распознавания); указать, какие поля будут подаваться оператору на верификацию и т.д.
- можно задать правила проверки значений полей. Правила помогают отделить документы, значения в которых не удовлетворяют тем или иным условиям, например, значение поля не соответствует значениям из определенной базы данных.
- задать способ экспорта данных. Данные могут экспортироваться в файл, базу данных, в Microsoft SharePoint, или согласно скриптовой процедуре.

После того как определение будет создано, его нужно опубликовать, чтобы оно стало доступно для распознавания документов.

Основные действия по созданию и редактированию определения документа производятся в окне **редактора определения документа**, которое открывается после завершения работы мастера создания нового определения. Для того чтобы открыть редактор определения документа из главного окна, нужно выбрать пункт меню **Проект > Определения документов...**, выделить название определения и нажать кнопку **Правка....**

Для создания нового определения выберите пункт меню **Проект > Определения документов...** и в появившемся диалоге нажмите кнопку **Новый...** Откроется Мастер создания нового определения. В диалоге **Создание нового определения документа** вы можете задать основные свойства определения: имя и комментарий, используемый язык и стиль написания. Задайте тип текста: выберите ICR (рукопечатный), если большинство полей документов будут заполнены от руки; выберите OCR (печатный), если значения в большинстве полей будут напечатаны, выберите тип печати в выпадающем списке. Позже, вы сможете поменять тип текста для отдельных полей, а указанный на данном этапе тип будет использоваться по умолчанию.

На втором шаге нужно загрузить или отсканировать изображение, по которому строится определение документа. (Если ваш документ состоит из нескольких страниц, то загрузите первую страницу, а для добавления остальных руководствуйтесь рекомендациями раздела «Как создать определение для многостраничных документов»). Вы можете отсканировать изображение страницы (желательно незаполненной) или загрузить его из файла. Если вы собираетесь обрабатывать полуструктурированные документы, то вам нужно использовать гибкое описание для создания описания документа. В этом случае отметьте опцию **Загрузить гибкое описание** и введите путь к файлу, содержащему гибкое описание, в формате *.afl (создается в программе ABBYY FlexiLayout Studio). Если создается гибкое определение документа, то этот шаг является последним.


На следующем шаге для жесткого определения документа выберите типы полей, которые должны быть найдены на изображении автоматически. Вы можете отметить метки и текстовые поля ввода, особенно хорошо производится автоматический поиск текстовых полей с разметкой и меток прямоугольной формы. Если же поля ввода на вашей форме не имеют разметки, а метки будут проставляться на белом фоне, то рекомендуется создать их вручную.

Если на форме есть реперы (статические элементы заданной формы, например, квадраты, использующиеся для идентификации и / или наложения определения документа), то они также будут найдены и отмечены на изображении автоматически.



3.1. Создание жесткой разметки

3.1.1. Геометрическая разметка объектов определения

После прохождения всех шагов мастера создания определения документа в окне редактора определения документа появится загруженное изображение, и на нем уже будут размечены поля тех типов, которые вы выбрали на последнем шаге мастера, а также реперы.

Автоматическое выделение объектов можно произвести и позднее, выбрав инструмент  и щелкнув кнопкой мыши в области элемента, который нужно выделить. Тип и расположение элемента будут определены автоматически.

Редактор определения документа снабжен набором удобных и наглядных инструментов для разметки полей и статических элементов. Редактор может работать в следующих режимах:

- режим разметки полей (инструмент ) и
- режим статических элементов (инструмент ).

Для того чтобы создать статический элемент или поле вручную, нужно выбрать режим разметки полей или статических элементов, нажать соответствующую кнопку на панели инструментов, а затем, нажав левую кнопку мыши, выделить прямоугольником нужный элемент. Либо выбрать инструмент и, удерживая нажатой клавишу Shift, щелкнуть левой кнопкой мыши в области объекта. В этом случае область поля или статического элемента будет определена автоматически.

Далее приведен список инструментов для создания элементов разного типа:

Поля:



- поле для ввода текста

Статические элементы:



- репер



- метка



- разделитель



- группа меток



- статический текст



- штрих-код



- штрих-код



- графический элемент (картинка)



- таблица



- группа полей

Штрих-код может быть как распознаваемым полем, так и статическим элементом. Надо внимательно выбирать подходящий режим в зависимости от назначения штрих-кода: если из штрих-кода будет извлечена информация, то отмечайте его область в режиме разметки полей, а если штрих-код используется для идентификации или наложения определения, отметьте его в режиме статических элементов.

Созданные поля появляются в списке на закладке **Поля** окна **Структура документа**. По умолчанию им присваиваются названия, совпадающие с пояснительным текстом или заголовком поля. Вы можете изменить название поля, например, выделив его в структуре документа и нажав F2. Если вы хотите присвоить полю название, совпадающее с пояснительным текстом, выделите поле, нажмите на правую кнопку мыши и выберите пункт локального меню **Распознать название с изображения...**, после чего нужно обвести текст с помощью курсора мыши.

На тип поля указывает пиктограмма в списке полей, а также цвет рамки на изображении. Статические элементы в списке не отображаются.


Объекты можно копировать (в том числе, в другие разделы документа), удалять, перемещать, изменять их размеры. Названия скопированных полей отличаются цифрами.

Можно выделить одновременно несколько объектов, удерживая нажатой кнопку Ctrl. Тогда выполняемое действие будет применено ко всем выделенным объектам. Для выделения объектов

используйте инструмент .

3.1.1.1. Группы полей

Поля могут быть объединены в группы для наглядности структуры документов, а также для создания повторяющихся групп полей. Например, город, улица и номер дома могут быть объединены в группу "Домашний адрес". Создать поля для "Рабочего адреса" можно простым копированием группы полей "Домашний адрес".

Для объединения полей в группу используйте инструмент .

Если в документах есть повторяющиеся группы полей, то после создания одной группы вы можете создать несколько экземпляров группы. Все свойства полей, а также правила, заданные внутри группы, будут работать для каждого экземпляра группы. См. «[Поля с несколькими экземплярами](#)».

Вы также можете копировать группу полей, но в этом случае результатом копирования будет создание еще одной независимой группы.


3.1.1.2. Поля без разметки

Поля могут не иметь соответствующей области на изображении. Поля без разметки в списке помечены красной звездочкой. Такие поля могут использоваться для хранения результатов вычислений, производимых со значениями распознаваемых полей.

Поля без разметки обладают всеми свойствами, присущими тому типу, к которому они относятся: они также могут подаваться на верификацию оператору, может производиться проверка формата, значения этих полей также можно экспортировать.

Создать поле без разметки можно двумя способами:

1. В окне редактора описания документа выберите пункт меню **Правка > Создать поле** и создайте поле нужного типа. Поле появится в списке и будет помечено звездочкой. В этом случае создается поле структуры документа, но не создается его область на изображении.
2. Второй способ заключается в удалении разметки обычного поля. Выделите нужное поле на изображении или в списке и выберите пункт локального меню **Удалить область**. Разметка будет удалена, а поле будет помечено красной звездочкой.


Чтобы создать область на изображении для полей без разметки, выберите  на панели инструментов и выделите мышью соответствующую область. Если в списке есть поля, помеченные звездочкой, то программа предложит вам выбрать название одного из таких полей. (Например, когда вы сначала задаете список полей документа, а потом определяете их местоположение.)



3.1.1.3. Разметка таблиц

ABBYY FlexiCapture 9.0 предоставляет возможность работы с таблицами. Для этого используется поле **Таблица**.

Для разметки таблиц в жестком описании документа предусмотрен ряд инструментов, с помощью которых вы сможете нарисовать таблицу, разметить строки и обозначить столбцы.

Нарисуйте область таблицы, выбрав инструмент . Шапка таблицы не должна быть включена в

эту область. Разметьте ячейки таблицы, используя для добавления разделителей инструмент . Для создания вертикальных разделителей поместите курсор в область таблицы. Перетащите пунктирный разделитель в нужное место таблицы и зафиксируйте его там щелчком мыши. Горизонтальные разделители создаются аналогично при нажатой клавише Alt. Также можно осуществить автоматический поиск разделителей. Для этого выделите созданную таблицу и воспользуйтесь пунктами локального меню: **Найти вертикальные разделители** и **Найти горизонтальные разделители**. Удалить любой из разделителей можно с помощью

инструмента . После добавления нужного количества разделителей обозначьте столбцы таблицы, выбрав инструмент  и указав курсором мыши область столбца. Каждый столбец содержит ячейки одного типа: текст, метки, графические элементы или штрих-коды. Программа предложит вам выбрать тип столбца при его обозначении.

После того, как будет выполнена геометрическая разметка таблицы, нужно задать свойства распознавания, верификации и тип данных для каждого столбца. Свойства столбцов задаются аналогично свойствам обычных полей документа.

3.1.1.4. Поля с несколькими областями

Если на вашей форме есть поля, область которых состоит из нескольких частей, например, таблицы, которые начинаются на одной странице, а продолжаются на следующей, то для их обработки можно создавать поля с несколькими областями.

Значения из всех областей одного поля объединяются и экспортируются вместе, как одно поле, в качестве разделителя используется перенос строки.

Чтобы создать поле с несколькими областями, создайте одну из областей поля, выделите ее, щелкните правой кнопкой мыши и выберите пункт локального меню **Создать продолжение области...** Выделите на странице место, которое будет продолжением созданной области. Повторите процедуру необходимое число раз.

3.1.1.5. Поля с несколькими экземплярами

В ваших документах могут встречаться повторяющиеся объекты – поля или группы полей, встречающиеся в документе несколько раз и описывающие схожие объекты, например, одни и те же сведения о нескольких сотрудниках, о детях, о счетах. Для таких объектов создаются поля с несколькими экземплярами.

Любое поле кроме таблицы может иметь несколько экземпляров, их области могут располагаться на расстоянии друг от друга, даже на разных страницах. Экземпляры полей обладают одними и теми же свойствами. Поля с несколькими экземплярами экспортируются в отдельные файлы или таблицы базы данных.

Поля с несколькими экземплярами удобно использовать для создания повторяющихся групп полей: создается группа полей, а затем для нее создается нужное количество экземпляров.

Чтобы создать поле с несколькими экземплярами, создайте один экземпляр, выделите его, щелкните правой кнопкой мыши в области поля и выберите пункт локального меню **Создать экземпляр области...** Создайте необходимое количество экземпляров области. Переместите их изображения в нужное место на странице.

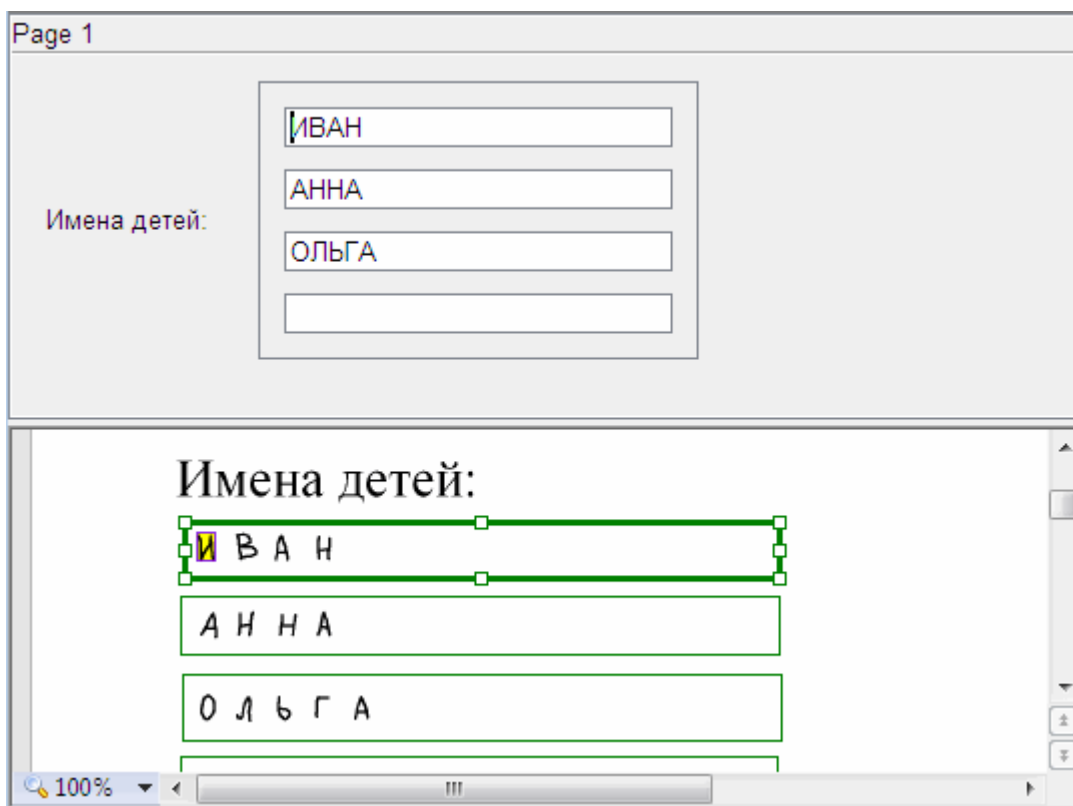



Рисунок 1. Пример использования полей с несколькими экземплярами

3.1.1.6. Исключение региона из распознавания

Иногда бывает необходимо исключить из распознавания некоторую область, например, область с пояснительным текстом, который мешает извлечению данных из поля (Рисунок 2). Чтобы исключить регион из распознавания, выберите инструмент  и выделите с помощью мыши область, которую нужно исключить.

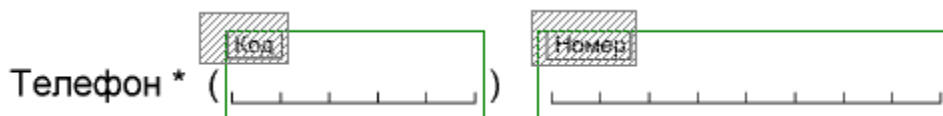



Рисунок 2. Пример использования нераспознаваемого региона

3.1.1.7. Удаление полей

Для того чтобы удалить поле, выделите его и нажмите клавишу **Delete** или выберите пункт **Удалить** локального меню. Для того чтобы удалить только разметку, а поле оставить в структуре документа, нужно нажать клавиши **Shift+Delete** или выбрать пункт **Удалить область** локального меню поля.

3.1.2. Статические элементы

Статические элементы – это объекты, из которых не извлекается информация. Они служат для наложения определения документа и идентификации. Примером статического элемента является репер.

Работа со статическими элементами производится в режиме, для переключения в который нужно нажать кнопку , только в этом режиме отображается разметка статических элементов.

Для наложения определения документа можно использовать все типы статических элементов, но лучшего результата можно добиться, если на документах есть специальные элементы – стандартные реперы: черные квадраты, кресты или уголки. Их следует отметить автоматически или вручную как статический элемент типа *Репер* и задать их форму в свойствах элемента на закладке **Общие**.

В качестве идентификаторов также можно использовать статические элементы любого типа. Программа определяет, к какому документу относится данная страница, по расположению идентификаторов или по их значению. Если в качестве идентификатора используется штрих-код, то можно задать его значение, это помогает точно и быстро идентифицировать страницу.

Чтобы использовать статический элемент для наложения определения документа и/или для идентификации, нужно отметить соответствующую опцию в диалоге свойств на закладке **Общие**. Диалог свойств любого элемента открывается при выборе пункта **Свойства...** локального меню элемента.

Статические элементы можно использовать при наложении определения документа (для этого отметьте опцию **Использовать для наложения определения документа**) и / или для идентификации определения документа (опция **Использовать для идентификации определения документа**).

Репер: задайте тип репера (квадрат, крест, уголок, прямоугольник). Для стандартных реперов рекомендуется отметить опцию **Использовать для наложения определения документа**. Опцию **Использовать для идентификации определения документа** есть смысл отмечать только в том случае, если расположение реперов уникально для данного раздела документа.

Статический текст: если статический текст будет использован для идентификации, вы можете ввести значение текста. Задавать значение текста нужно только в том случае, если страницу нельзя идентифицировать по расположению этого текста, а только по его значению. (Например, страницы отличаются только текстом заголовков, а расположение и размер заголовков одинаковое).

Разделитель: выберите, будет ли использоваться разделитель для идентификации и наложения определения.

Штрих-код: если штрих-код будет использоваться в качестве идентификатора, вы можете ввести значение штрих-кода. На закладке **Распознавание** введите тип штрих-кода, ориентацию и опции обработки изображения.

3.1.2.1. Особенности штрих-кода

Если штрих-код используется в качестве идентификатора, то это реперный штрих-код и относится он к статическим элементам. Создавать его следует в режиме работы со статическими элементами. Диалог свойств для такого штрих-кода содержит две закладки: **Общие** и **Распознавание**.

Если же вы планируете использовать штрих-код для извлечения информации, то такой штрих-код является полем, и создавать его следует в режиме работы с полями. Диалог свойств содержит все закладки, присущие полям, а именно, **Общие**, **Тип данных**, **Распознавание**, **Верификация**, **Правила**. Значение такого штрих-кода будет распознано и, в зависимости от настроек, подано на верификацию и экспортировано.

3.1.3. Свойства полей

Очень важно правильно задать свойства полей. Они влияют на качество распознавания полей, и на то, будут ли их значения экспортироваться и подаваться оператору на проверку. Некоторые свойства очень важны для распознавания данных, например, способ разметки для текстового поля ввода: важно правильно задать это свойство, чтобы разметка, не исчезающая при сканировании, была отделена от символов. Тогда результатом распознавания будет только внесенный в поле текст, без лишних элементов.

Правильно заданные свойства полей приводят к хорошему результату распознавания данных, а это значительно снижает трудозатраты операторов. Задавая свойства полей, верификацию документов можно свести к минимуму.

Когда вы создали поле определенного типа, ему присваиваются свойства по умолчанию. Некоторые свойства по умолчанию наследуются из определения документа (например, язык), другие – определяются автоматически (например, тип разметки). Их можно изменить в диалоге, который открывается посредством выбора пункта **Свойства** контекстного меню любого объекта. Поле каждого типа имеет свой набор свойств.

3.1.3.1. Общие свойства полей

Закладка **Общие** (Рисунок 3) присутствует в диалоге свойств всех полей. На этой закладке задаются **Название поля** и его описание. Вы можете изменить название, которое задается автоматически при создании поля и соответствует пояснительному тексту, расположенному наиболее близко от поля. **Заголовок** – это название поля, которое отобразится в форме данных. Тип поля отображается на этой закладке в виде пиктограммы справа от имени.

Кроме того, на закладке **Общие** можно задать следующие опции:

- **Экспортировать значение поля** – отключите эту опцию, если вам не нужно экспортировать значение поля. Отключение может понадобиться, если, например, значение этого поля используется для получения значения вычисляемого поля (см. [Проверка с помощью правил](#)), и нужно экспортировать только конечный результат.
- **Только для чтения** – отметьте эту опцию, если поле должно быть доступно только для просмотра, то есть оператор не сможет изменить его значение в редакторе документа. Вы можете отметить эту опцию для полей, значение которых должно вычисляться автоматически, с помощью правила, например, для поля, в котором планируется сохранять сумму или объединение значений других полей (см. [Проверка с помощью правил](#)).
- **Показывать в Форме данных** – отключите эту опцию, если не хотите, чтобы поле присутствовало на форме данных при просмотре документа. Если в таком поле возникнет ошибка, то исправить ее в окне редактора документов будет невозможно (поскольку поле не будет отображаться), поэтому рекомендуется указывать для таких полей настройки, не вызывающие ошибок.
- **Не может быть пустым** – отметьте эту опцию, если поле должно быть обязательно заполнено. Если поле окажется пустым, появится ошибка формата.
- **Индексное поле** – отметьте эту опцию, если предполагается производить индексацию по этому полю. Тогда значение поля можно будет вывести для каждого документа в списке, и оператор сможет сортировать и искать документы по значению этого поля. Также значение индексного поля можно использовать при экспорте для именованя файлов.

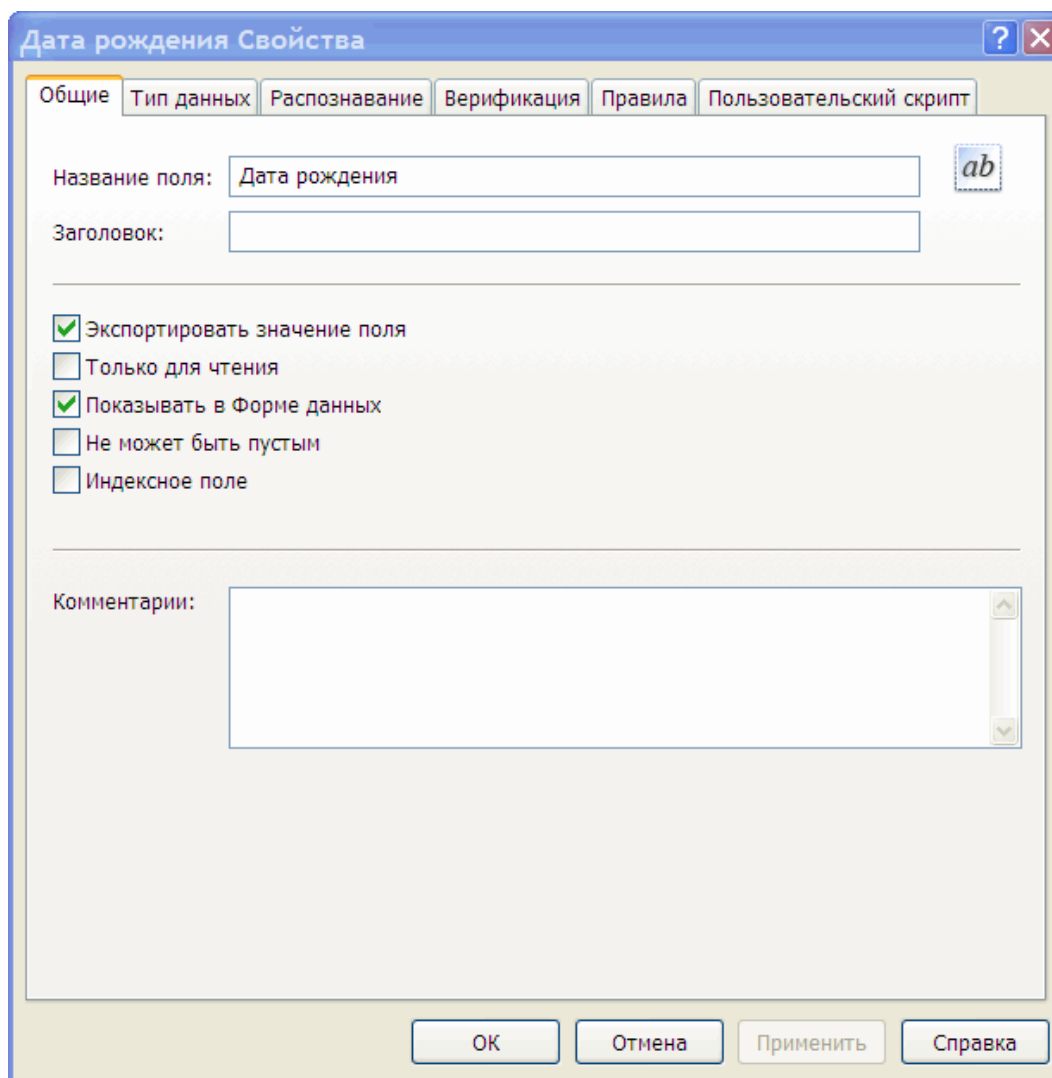


Рисунок 3. Диалог задания свойств поля, закладка для задания основных опций

3.1.3.2. Типы данных

Тип данных определяет множество возможных значений поля и его допустимый формат. Если значение, введенное в поле, не будет соответствовать заданному типу данных, то оператор получит сообщение об ошибке в процессе верификации поля. Тип данных для текста обычно имеет простое обозначение области допустимых значений – дата, время, адрес, ИНН, сумма денег, число. Для меток тип данных представляет собой значения, которые принимает поле в случае отмеченной и неотмеченной метки.

3.1.3.2.1. Типы данных текстового поля ввода

Для текстового поля ввода важно правильно задать **тип данных**, тем самым указывая программе, данные какого типа мы ожидаем увидеть в этом поле: будут ли это одни цифры, буквы какого-либо алфавита, символы из определенного набора, дата или что-то другое. Программа обладает гибким механизмом задания типов данных. Пользователю предоставляется готовый набор типов данных, который включает в себя наиболее часто используемые типы. Кроме того, пользователь может создавать свои типы данных, подходящие для решения конкретной задачи.

При задании типов можно настроить проверку формата введенного значения, а также задать ограничения значения, например, минимальное и максимальное значения для числа, промежутки времени для даты, максимальное количество символов для текстовых полей.

Тип данных задается на закладке **Тип данных** диалога свойств поля (Рисунок 4).

Выберите нужную категорию из списка **Содержимое**. В поле **Детали** вы увидите описание одного из типов данных этой категории (заданного по умолчанию или настроенного вами ранее).

Если отмечена опция **Обрабатывать как текст**, то значения полей с любым содержимым будут обрабатываться и экспортироваться как текстовые. Также не будет проводиться проверка на соответствие содержимого формату поля.

Если вам необходимо поменять **Язык** распознавания или задать более узкий тип данных, нажмите на кнопку **Правка...** справа от описания.

Для стандартных типов (в списке **Тип содержимого** выбрано **Стандартный**) можно

- для Текста - выбрать несколько языков распознавания (кнопка «...»). Использовать встроенный словарь и/или пользовательский словарь. Тогда при распознавании поля будет учитываться информация из словарей.
- для Числа - отметить опцию **Целое**, если предполагается, что значение поля будет целым числом;
- для Даты - выбрать порядок компонентов даты, допустить возможность написания месяца словом, содержания времени или дня недели;
- для Адреса, Имени, Кода - подключить пользовательский словарь.

Специальные типы (в списке **Тип содержимого** выбрано **Специальный**) содержат predefined типы данных, среди которых вы можете выбрать подходящий для данного поля. При выборе руководствуйтесь описанием, которое отображается в нижней части диалога.

Вы можете создать свой тип данных, если вам не подходит ни один тип из списка.

1. Для этого выберите одно из значений в списке **Содержимое** на закладке **Тип данных** свойств поля. Вы можете выбрать любое значение из списка, которое соответствует смыслу вашего пользовательского типа: создание нового типа от этого не зависит, но храниться новый тип данных будет в выбранной категории.
2. Затем нажмите кнопку **Правка...** справа от поля **Детали**. В открывшемся диалоге выберите значение **Специальный** в поле **Тип содержимого**. Нажмите кнопку **Новый...** справа от поля **Выберите тип данных**.
3. Следуйте инструкциям Мастера создания нового типа данных.

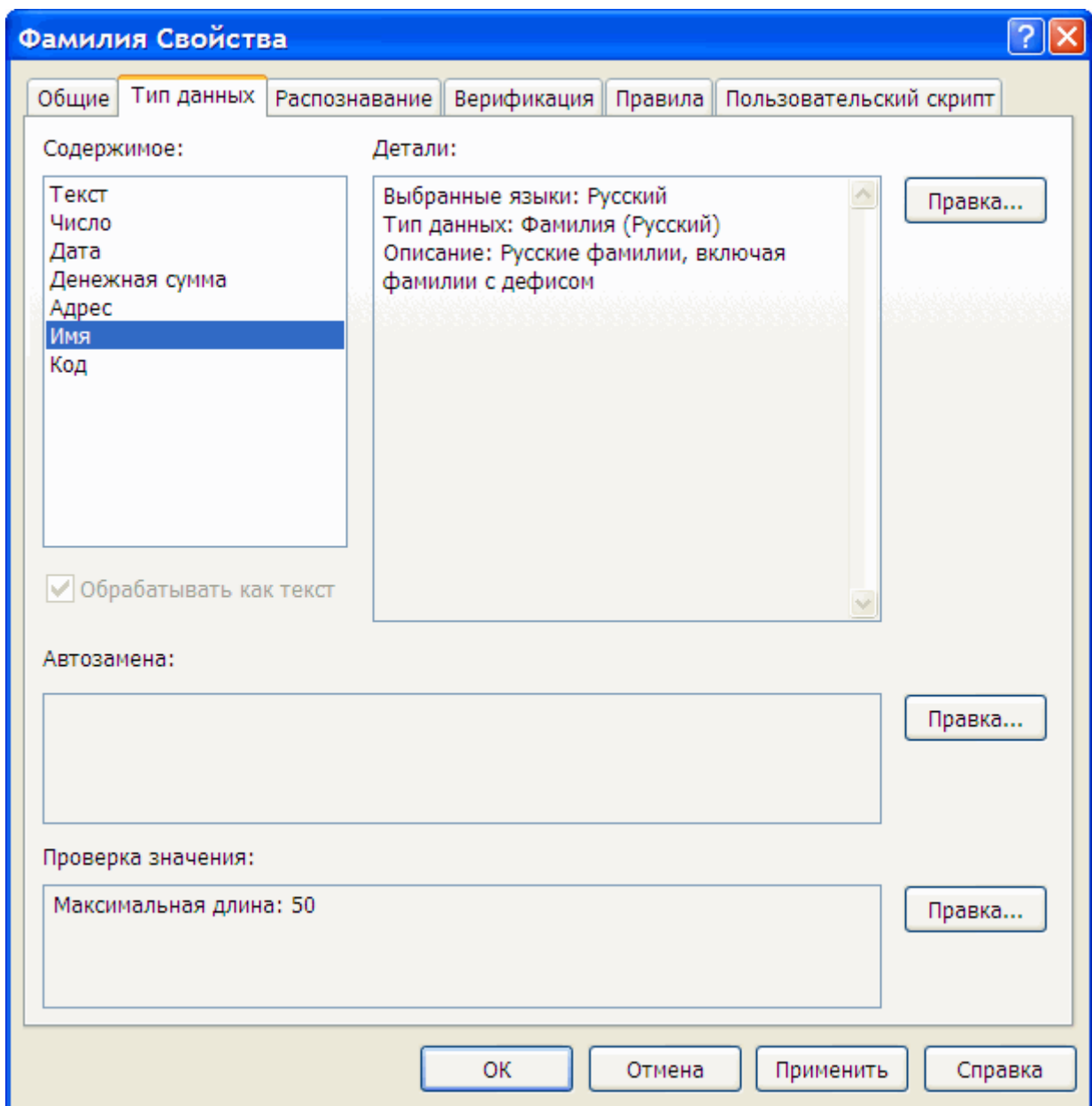


Рисунок 4. Диалог задания свойств поля, закладка для задания типов данных (для текстового поля ввода)

Для любых типов данных можно произвести автоматическую обработку распознанного значения: убрать лишние пробелы, преобразовать буквы в заглавные или строчные, автоматически заменить указанные символы или текст. Для этого нажмите кнопку **Правка...** справа от поля **Автозамена**. В открывшемся диалоге задайте необходимые параметры обработки текста (Рисунок 5).

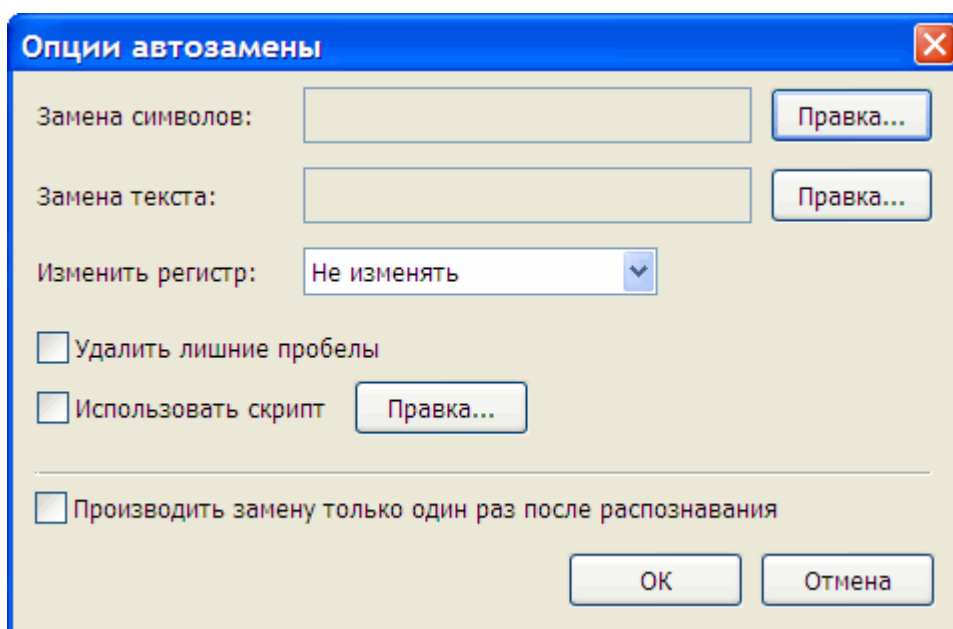


Рисунок 5. Диалог задания параметров автозамены

Также можно настроить проверку распознанного значения поля: проверять, удовлетворяет ли значение заданным условиям (например, для даты – попадает ли значение в определенный интервал, для текстовых полей – является ли значение допустимым или соответствует ли оно определенному формату и т. д.) Чтобы задать ограничение области значений для поля, нажмите кнопку **Правка...** справа от поля **Проверка значения**. (Рисунок 6).

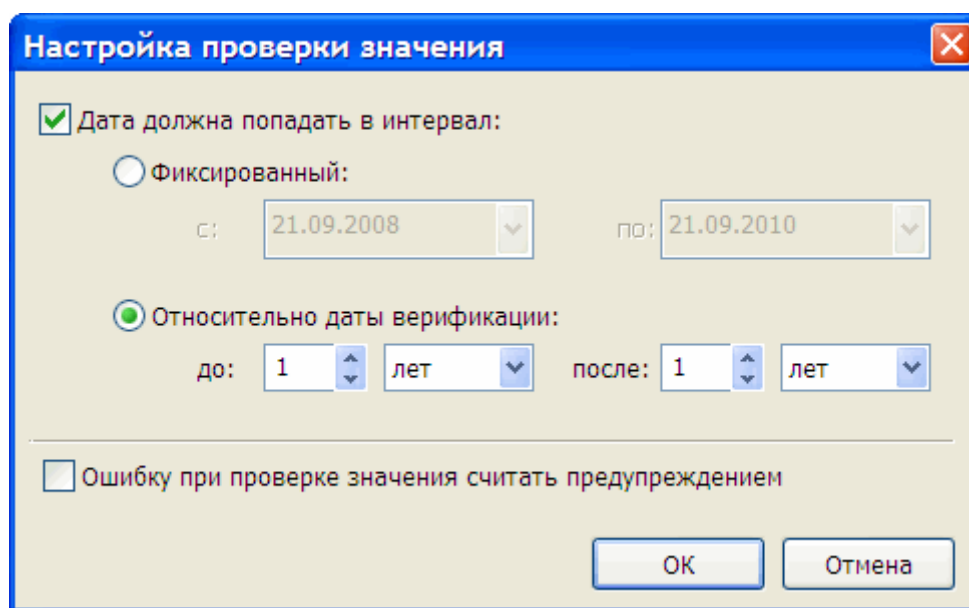


Рисунок 6. Диалог настройки проверки значения

3.1.3.2.2. Типы данных для меток

Для меток на закладке **Тип данных** определяются значения, которые будут присваиваться полям в случае отмеченных и неотмеченных меток (Рисунок 7).

Полям меток могут присваиваться следующие значения:

- *Да* – если метка отмечена, *Нет* – если метка не отмечена;

- 1 – если метка отмечена, 0 – если метка не отмечена;
- *Название метки* – если метка отмечена, *Пустая строка* – если метка не отмечена;
- вы также можете отметить опцию *Пользовательские значения* и ввести свои значения.

Замечание. Если метки объединены в группу, то они обладают общими свойствами. Свойства задаются только для всей группы меток, значения меток также задаются в свойствах группы.

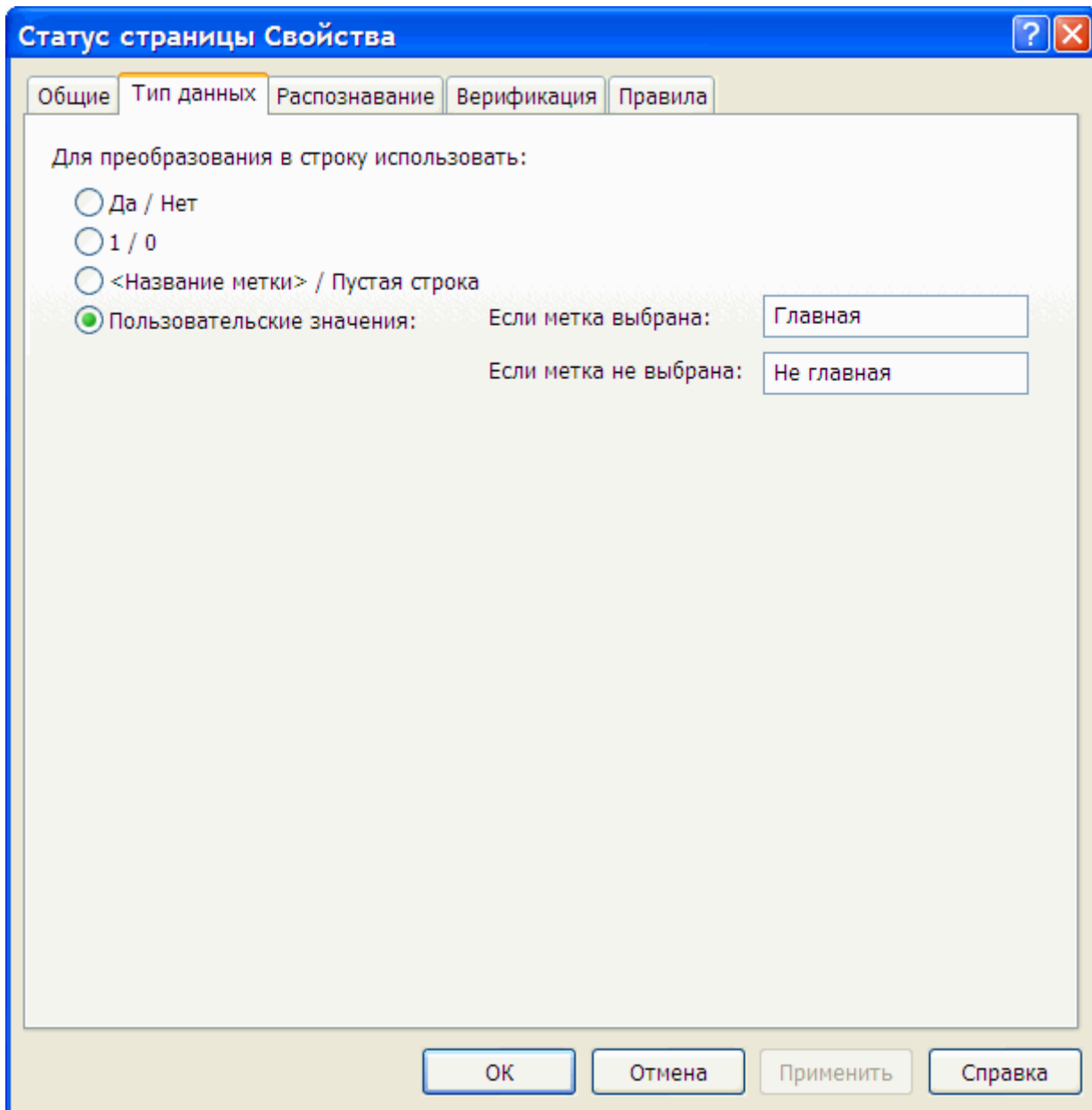


Рисунок 7. Диалог задания свойств поля, закладка для задания типов данных (для метки, не входящей в группу)

3.1.3.2.3. Типы данных для групп меток

В диалоге свойств групп меток на закладке **Тип данных** отображается список названий меток, входящих в данную группу (Рисунок 8).

Снимите отметку с опции **Допускать, если не выбрано ни одной метки**, если в группе должна быть обязательно отмечена хотя бы одна метка.

Если в данной группе можно отмечать несколько меток, то отметьте опцию **Допускать множественный выбор**.

Вы можете определить экспортируемые значения в случае, если не было выбрано ни одной метки и в случае, если в группе выбрано более одной метки. Для этого выберите в списке *<Ничего не выбрано>* или *<Множественный выбор>*, нажмите на кнопку **Правка...** и в открывшемся диалоге задайте значение в поле **Экспортируемое значение**. Если значения для экспорта не заданы, то в случае, если ни одна метка не выбрана, экспортируется пустая строка, а в случае выбора нескольких меток экспортируются их значения через запятую.

Если отмечена опция **Ошибку при проверке считать предупреждением**, то вместо сообщения об ошибке выдается предупреждение.

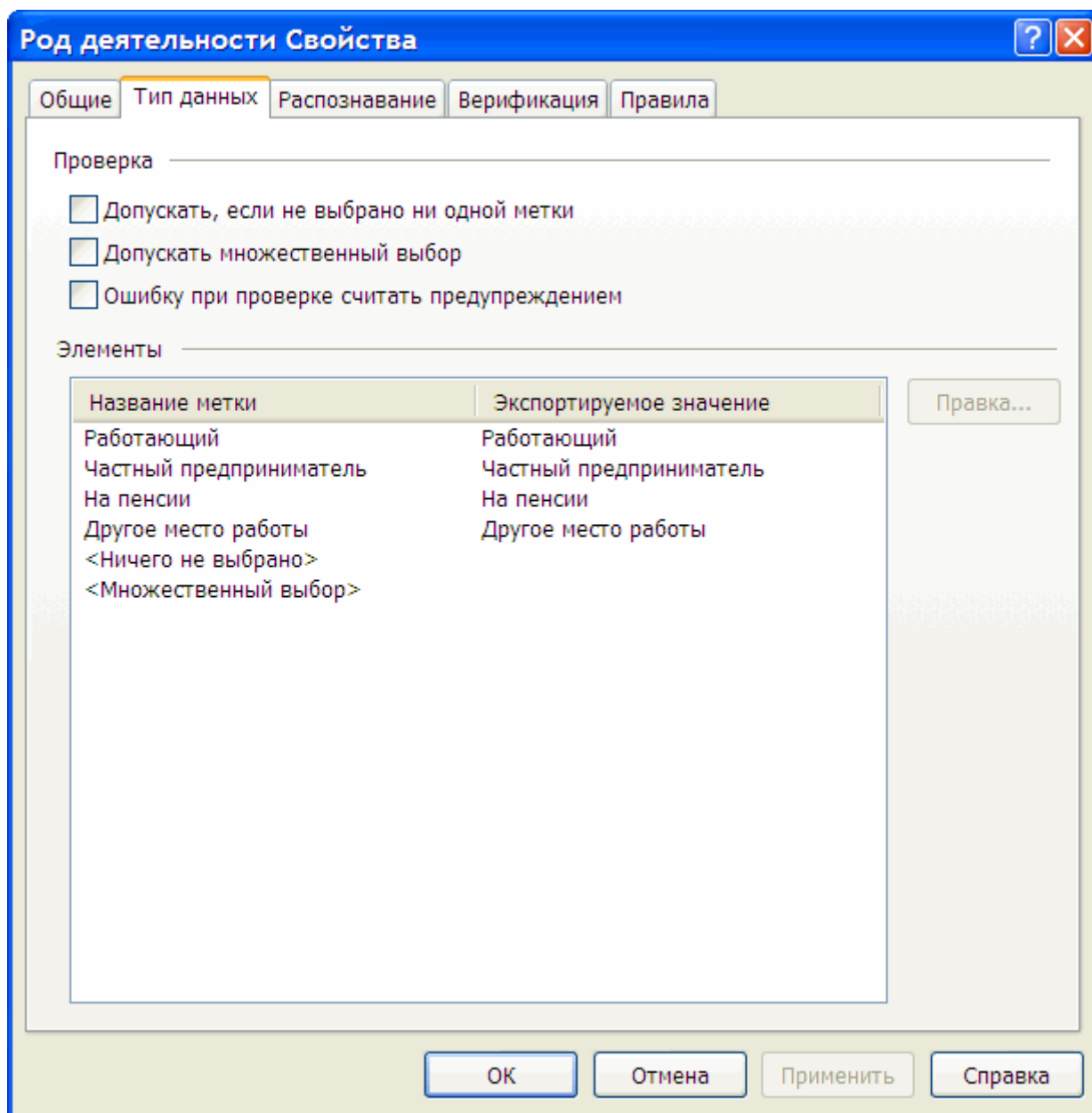


Рисунок 8. Диалог задания свойств поля, закладка для задания типов данных (для групп меток)

3.1.3.3. Свойства распознавания полей

В программе ABBYY FlexiCapture 9.0 есть возможность задать опции распознавания для каждого поля. Правильно определенные свойства поля на закладке **Распознавание** диалога задания свойств значительно повысят качество распознавания и снизят вероятность появления ошибок. Для разных типов поля эти свойства отличаются, рассмотрим подробнее свойства распознавания в зависимости от типа поля.

3.1.3.3.1. Свойства распознавания текстового поля ввода

Опцию **Не распознавать значение поля, оно будет введено вручную** следует отметить, если это поле по каким-либо причинам не требуется распознавать, и его значение будет вводиться вручную оператором (например, если значение поля сложно распознать из-за слитного написания букв). В этом случае не нужно задавать другие свойства распознавания, так как такое поле распознаваться не будет, а при проверке оператору будет предложено ввести значение поля.

Выберите тип текста в зависимости от того, будет ли он написан от руки – **ICR (рукопечатный)**, или напечатан – **OCR (печатный)**. Для напечатанного текста выберите стиль печати (типографским способом, на матричном принтере, печатной машинке и т.д.) Для того чтобы задать несколько типов текста или использовать шаблон, отметьте опцию **Дополнительно** и нажмите **Изменить...**

Выберите **Тип разметки**, выбор облегчают образцы разметки в выпадающем списке. Если разметка исчезает при сканировании, то рекомендуется выбрать тип разметки для моноширинного текста (*Серые прямоугольники*), а если разметка не исчезает при сканировании и разделена на ячейки для знакомест, то вам обязательно нужно ввести общее число ячеек (число ячеек может быть определено автоматически). Тип *Простой текст* используется для полей без разметки, как правило, для текста, напечатанного типографским способом.

Вы можете выбрать **Регистр** букв, чтобы при распознавании осуществлялся поиск букв определенного регистра. Если поле может быть заполнено как строчными, так и заглавными буквами, оставьте значение **Любой**.

Выберите горизонтальную или вертикальную **Ориентацию** текста.

Для однострочного поля отметьте опцию **Однострочное**, а для поля, значение которого всегда будет состоять из одного слова (т.е. не содержит пробелов) – опцию **Одно слово**.

Задайте опции обработки изображения при распознавании: изображение поля может быть очищено от «мусора», причем можно изменять размер удаляемых точек; изображение может быть инвертировано; также возможно удаление текстуры.

Замечание. По умолчанию, если отмечена опция **Очищать изображение**, размер «мусора» определяется автоматически. Если вы хотите задать свой размер, то отметьте опцию **Чистить мусор только определенного размера** и введите размер.

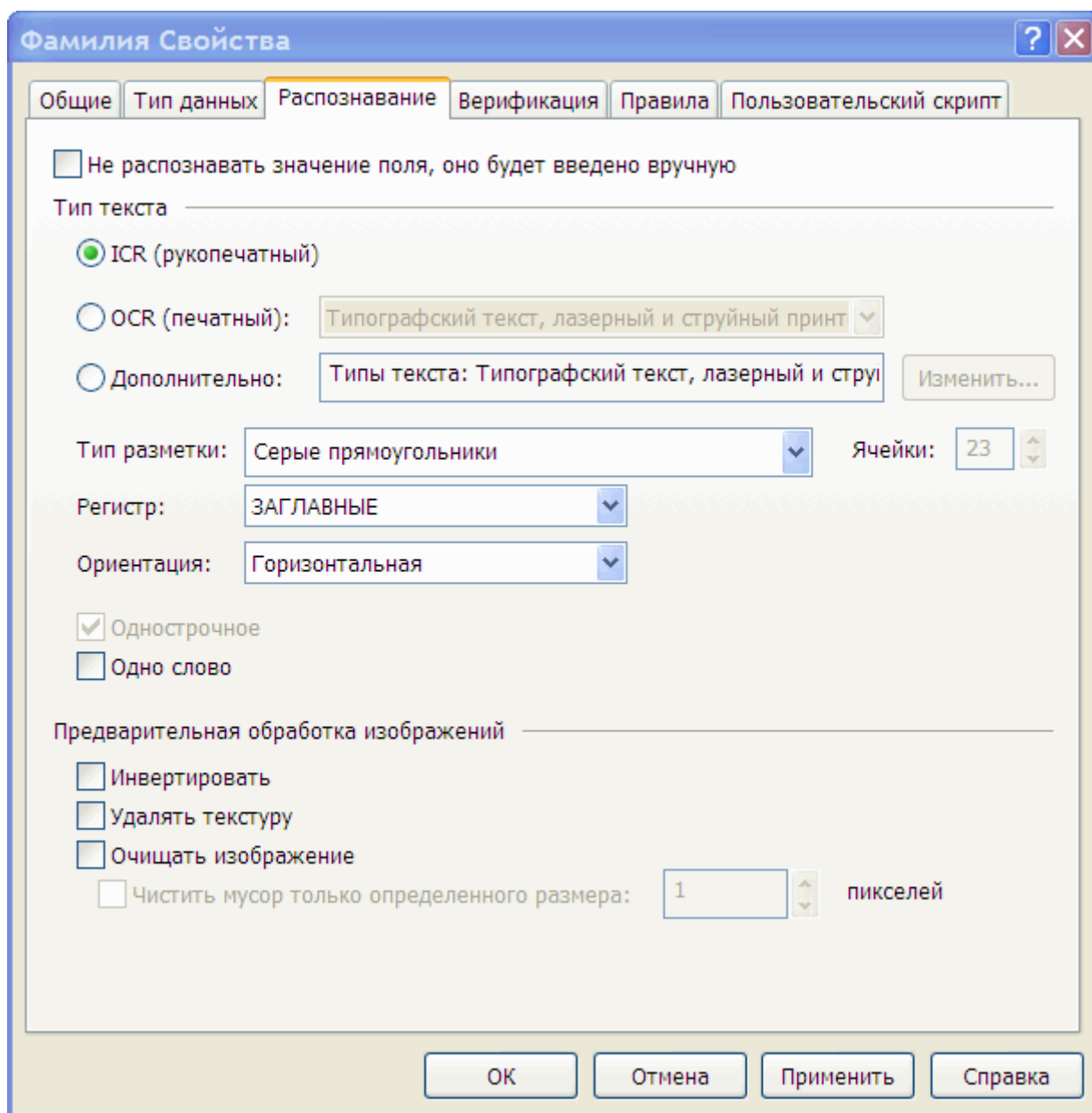


Рисунок 9. Диалог задания свойств поля, закладка для задания опций распознавания (для текстового поля ввода)

3.1.3.3.2. Свойства распознавания меток и групп меток

Для правильного распознавания меток нужно задать способ распознавания метки **Тип метки**. Если метка ограничена квадратом, задайте для нее тип **Квадрат**; если метка ставится на белом фоне без ограничителя (или с ограничителем, который исчезает при сканировании), задайте тип **Без рамки**. Программа будет определять форму метки автоматически, если тип метки **Автоматическое определение**. (В этом случае изображение метки на бланке не должно быть заполнено, т.к. программа будет определять проставлена ли метка, сравнивая изображение данной области в обрабатываемом документе и на бланке, на основе которого создается описание документа).

Вы можете допустить исправления для каких-либо меток: в случае ошибочного проставления метки, заполняющий может заштриховать всю клетку. Полностью заштрихованные метки в этом

случае будут считаться неотмеченными. Но если вы выбрали тип **Автоматическое определение**, то для таких меток исправления допускать нельзя.

Для меток можно задать опции обработки изображения также как и для текстового поля.

Если метки объединены в группу, то они обладают общими свойствами. Свойства распознавания задаются аналогично, но для всей группы меток.

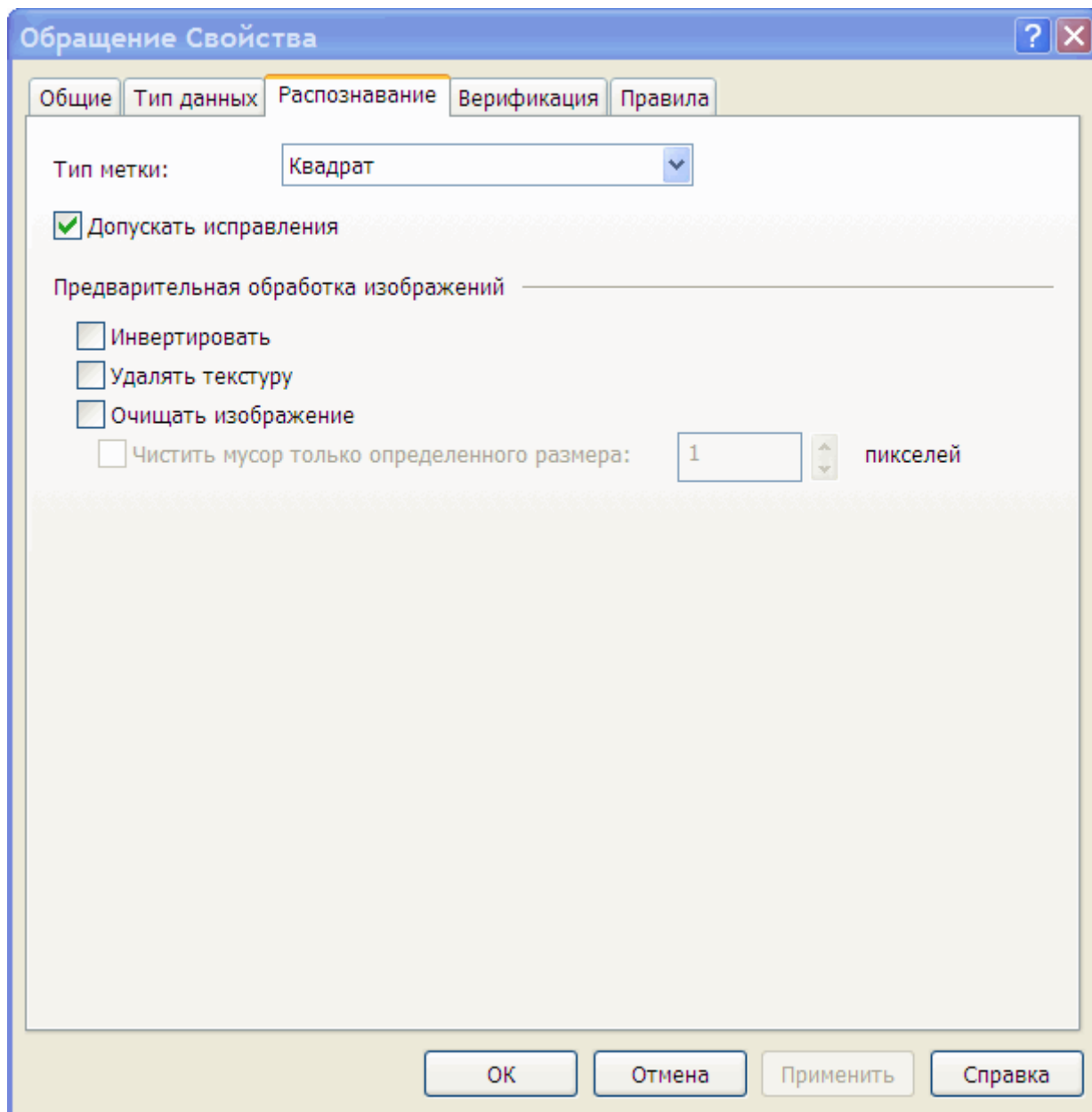


Рисунок 10. Диалог задания свойств поля, закладка для задания опций распознавания (для метки)

3.1.3.3.3. Свойства распознавания штрих-кода

Свойства распознавания штрих-кода, являющегося полем, совпадают со свойствами штрих-кода статического элемента, для поля также задается тип штрих-кода, ориентация и опции очистки изображения. Единственное отличие в том, что значение поля может вводиться оператором

вручную, для этого нужно отметить опцию **Не распознавать значение поля, оно будет введено вручную**.

3.1.3.3.4. Свойства распознавания картинки

На закладке **Распознавание** для поля картинка отметьте опцию **Исключить из распознавания**, если область картинки должна быть исключена из области распознавания остальных полей. (Эта опция доступна только для поля Картинка и используется для совместимости с шаблонами FormReader 6.5.)

3.1.3.4. Параметры верификации

Верификация – это проверка распознанных данных оператором. В процессе создания описания документа вы можете задать опции верификации на закладке **Верификация** диалога задания свойств поля (Рисунок 11). Неуверенно распознанные символы помечаются программой и подаются оператору на проверку, но можно задать свойства поля таким образом, что поле будет подаваться на проверку, даже если в нем нет неуверенно распознанных символов, либо наоборот, не будет подаваться на проверку, даже при наличии неуверенно распознанных символов. Значение *все* (обязательная верификация) можно выбрать в том случае, если для вас недопустима ошибка в значении данного поля.

Включение символов поля в групповую и контекстную верификацию осуществляется с помощью соответствующих опций. В групповой верификации верифицируются только метки и цифры. Если в поле могут быть другие символы (буквы), то его надо включать в контекстную верификацию.

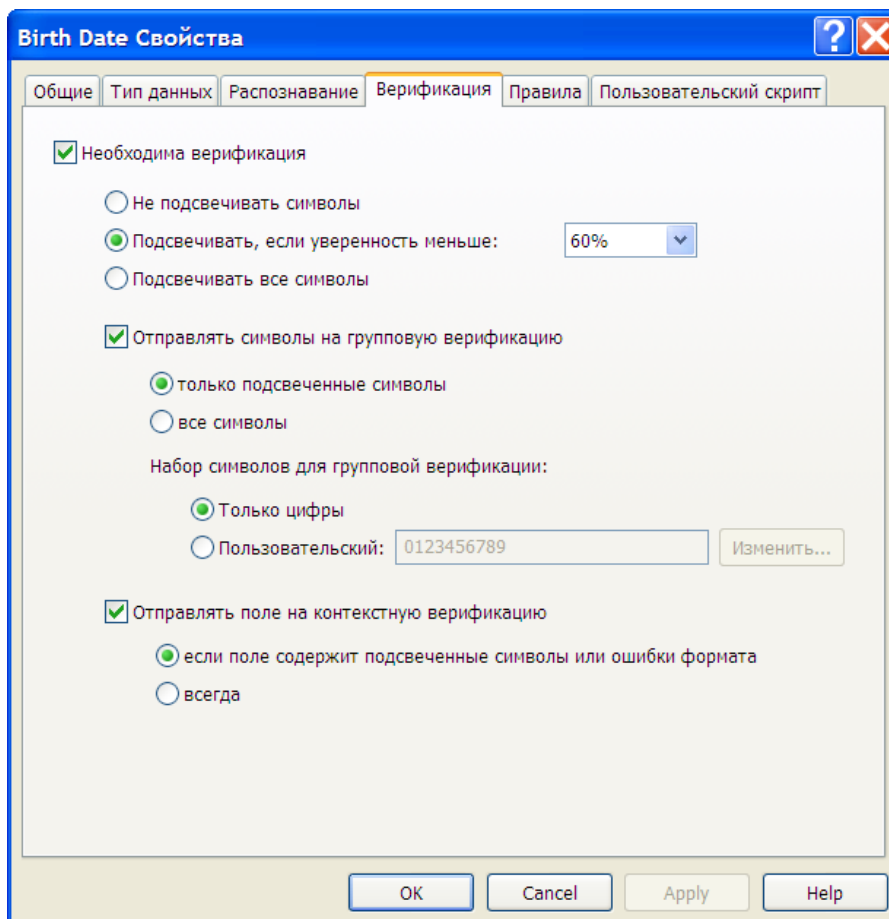


Рисунок 11. Диалог задания свойств поля, закладка для задания опций верификации

3.1.3.5. Параметры экспорта картинок

Для поля **Картинка** в описании документа можно настроить параметры экспорта, такие как тип файла, качество, цвет и разрешение. Эти параметры задаются на закладке **Экспорт** диалога задания свойств поля (пункт **Свойства...** локального меню поля). Можно задать:

- **Тип файла** (TIFF, JPEG, BMP, JPEG2000, PCX packbits, PNG).
- **Качество**. Для файлов формата TIFF, JPEG или JPEG2000 можно выбрать качество экспортируемого файла (наилучшее, высокое, обычное, низкое).
- **Тип цвета** - цветной, градации серого, полутона (ч/б).

Чтобы установить разрешение экспортируемой картинки, отметьте опцию **Установить разрешение равным** и введите разрешение или выберите нужное значение из списка.

3.1.3.6. Проверка с помощью правил

Правила служат для автоматической проверки распознанных данных. Правила, наряду с типами данных, позволяют задать ограничения на данные, то есть определить условия, которым должны удовлетворять значения в тех или иных полях. Если эти значения в заполненных документах не удовлетворяют условиям, то такие страницы помечаются флажком с соответствующим сообщением. Основная цель применения правил – это проверка целостности данных документа. Кроме того, с помощью правил можно обработать распознанные данные: объединить значения нескольких полей или заменить распознанные значения значениями из базы данных.

Правила выполняются в последнюю очередь, после автокоррекции значений, проверки данных на соответствие формату поля и валидации данных (проверка попадания даты в указанный диапазон, числа в определенный интервал и т.д.).

ABBYY FlexiCapture 9.0 позволяет задать правила следующих видов:

- **Проверка суммы** – проверка суммы значений нескольких полей. Сумма значений указанных полей сравнивается с заданным числом или со значением в другом поле. Например, если в вашем документе указан оклад сотрудника и премия, а также есть поле с общей суммой дохода, то мы можем задать правило, проверяющее совпадает ли сумма оклада и премии с общей суммой дохода. При несовпадении будет выдаваться сообщение об ошибке.
- **Сравнение значений полей** – сравнение значений нескольких полей (в том числе из разных разделов). Это правило можно применять, если в вашем документе есть несколько полей, значения которых должны быть одинаковыми. При несовпадении значений для документа будет выдаваться сообщение об ошибке правила.
- **Проверка по базе данных** – производится проверка введенных значений путем их сравнения со значениями из определенной базы данных.
- **Объединение значений полей** – слияние значений нескольких полей. Значения полей при слиянии можно отделять друг от друга точкой, пробелом или другими разделителями. Например, бывает удобно не распознавать всю дату целиком, а составлять ее из отдельно распознанных полей «день», «месяц» и «год», добавив точки в качестве разделителей. Результат слияния может храниться в любом поле описания документа, удобно использовать с этой целью поля без разметки. (См. [Поля без разметки](#)).

- Правило проверки суммы – сравнение суммы цифрами с суммой прописью (только на русском языке). Правило доступно только если для описания документа установлен русский язык распознавания.
- Скрипт – ограничения описываются пользователем с помощью скрипта. Подробно язык скриптов описан в справочной системе приложения. Поскольку порядок выполнения правил не указывается, поле не может быть доступно для записи из двух скриптовых правил. Только для чтения оно может быть доступно в любом количестве правил

Правила задаются на закладке **Правила** диалога задания свойств поля (Рисунок 12). Правила могут затрагивать значения одного или сразу нескольких полей.

Правилу можно присвоить статус ошибки или предупреждения, при невыполнении условия поле помечается красным флажком в случае ошибки, желтым – в случае предупреждения.

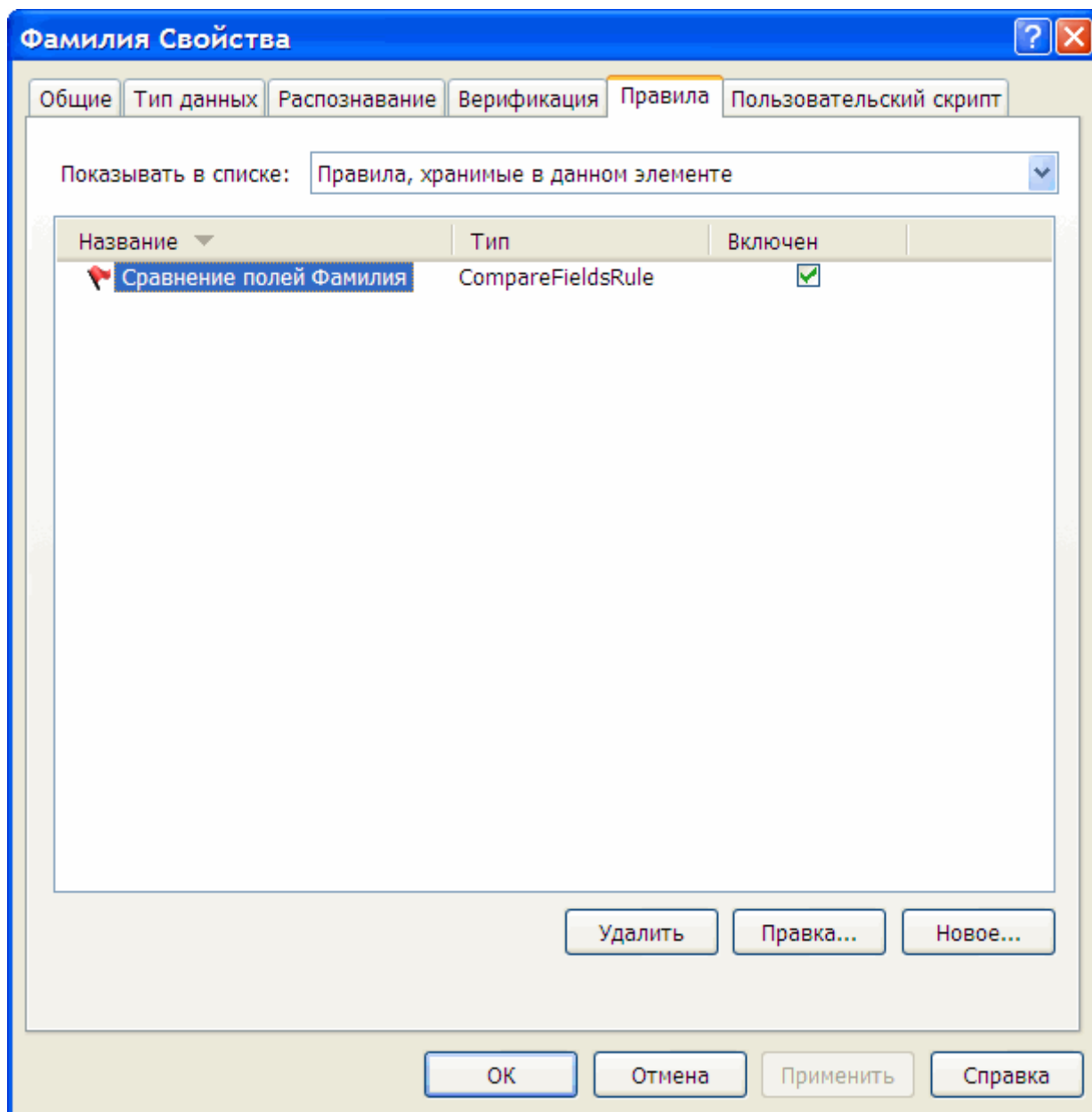


Рисунок 12. Диалог задания свойств поля, закладка для задания правил

3.1.3.7. Пользовательский скрипт

При необходимости для поля можно создать пользовательский скрипт. В отличие от скриптового правила, запускающегося автоматически, пользовательский скрипт выполняется по запросу пользователя при просмотре поля во время контекстной верификации либо при просмотре документа в Редакторе документа. Скрипт запускается с помощью кнопки ... справа от значения поля. Подробно язык скриптов описан в справочной системе приложения.

Пользовательский скрипт задается на закладке **Пользовательский скрипт** (Рисунок 13). Добавление и удаление полей, информация о которых будет доступна в пользовательском скрипте, осуществляется с помощью кнопок **Добавить...** и **Удалить** соответственно. Для перехода в окно редактирования скрипта необходимо нажать кнопку **Правка...** Для включения скрипта необходимо отметить опцию **Включить пользовательский скрипт**.

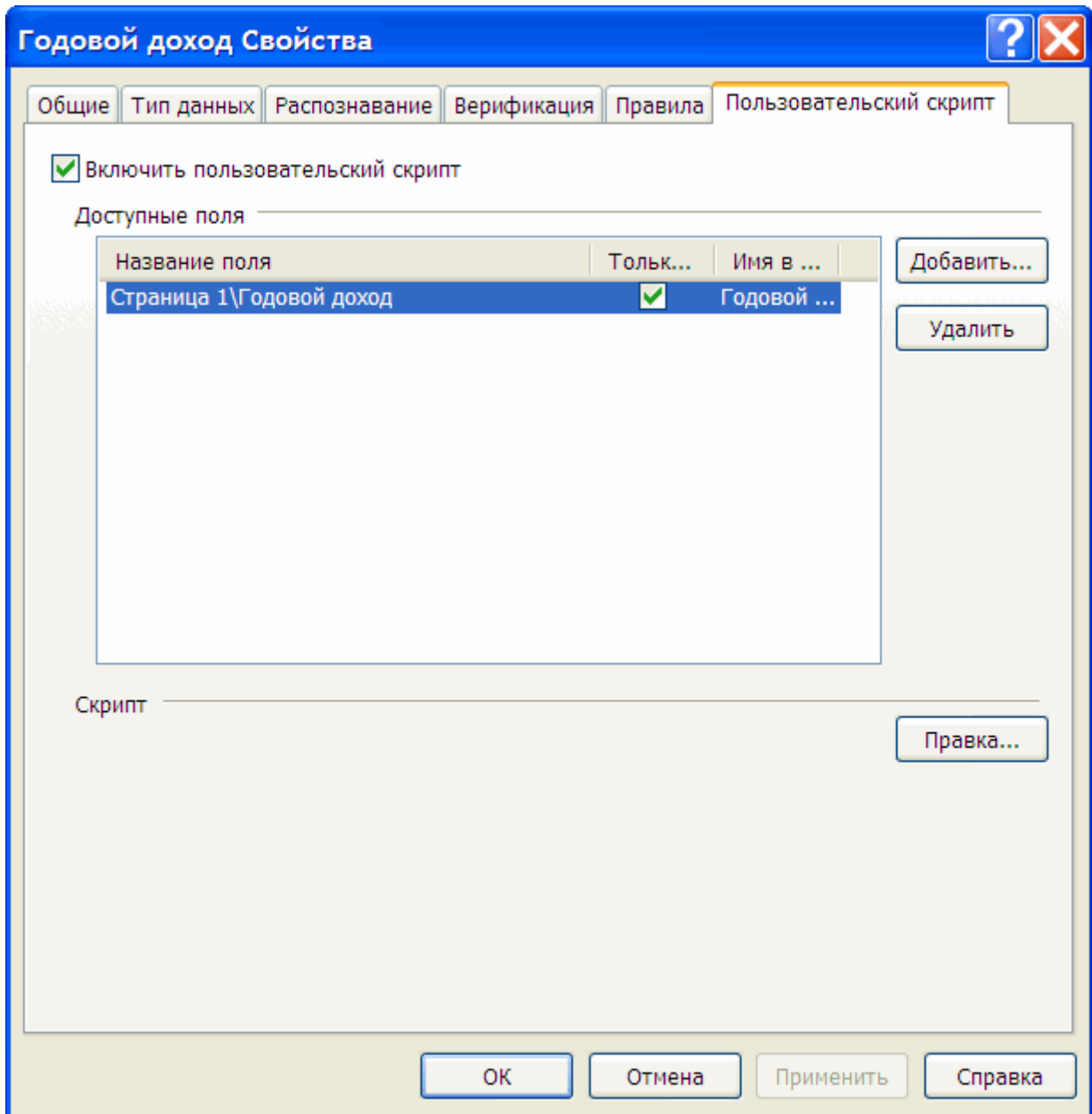


Рисунок 13. Диалог задания свойств поля, закладка для задания пользовательского скрипта

3.1.4. Как создать определение для многостраничных документов

ABBYY FlexiCapture 9.0 позволяет создавать определения многостраничных документов. Определение может состоять из произвольного количества **разделов**, каждый из которых может содержать одну или несколько **страниц**. Для определений многостраничных документов задается порядок следования разделов, их количество и правила сборки страниц в документы.

После завершения работы Мастера создания нового определения документа создается простое определение, не содержащее разделов. Вы можете добавлять в определение страницы. При добавлении страницы ее изображение добавляется в область изображения в окне редактора определений документов, а поля, созданные на ней, будут фигурировать в общем списке полей. В этом случае документ состоит из одного раздела, хотя он может содержать несколько страниц.

Добавлять страницы в один раздел следует, если в документе есть блоки, переходящие со страницы на страницу, повторяющиеся на разных страницах блоки, таблицы, переходящие на следующую страницу. Такой же способ используется при загрузке многостраничного гибкого описания (см. Настройка системы для ввода гибких документов). Все поля гибкого описания должны относиться к одному разделу определения документа.

Другой способ заключается в добавлении к определению разделов. В наиболее простом случае один раздел содержит одну страницу. Так можно поступать, если наборы полей на страницах независимы. Такой способ более наглядный: в структуре документа видно, к какому разделу относятся поля, а также вы можете задать произвольную структуру сборки документа. Например, в вашем документе 3 страницы, причем первая и третья содержатся в документе 1 раз, а вторая может повторяться от 2 до 5 раз. Тогда нужно создать для каждой страницы отдельный раздел, а потом задать структуру документа, то есть последовательность и количество повторений разделов.

Более сложный случай – определение, состоящее из нескольких разделов, причем некоторые разделы содержат более одной страницы. Это, например, определение, состоящее из жестких разделов и гибкого многостраничного раздела. Или определение, описывающее документы, содержащие двусторонний лист, который может повторяться несколько раз. Допустим, вам нужно создать определение документа, содержащего титульный лист (страницу А) и некоторое количество двусторонних листов (последовательность страниц ВСВС...). Тогда создаем раздел А, состоящий из одной страницы А и раздел ВС, состоящий из двух страниц В и С. Структуру документа зададим следующим образом: раздел А повторяется 1 раз, за ним следует раздел ВС, повторяющийся, например, от 3 до 7 раз. Аналогично возможно повторение страниц внутри раздела. Повторение надо использовать осторожно, поскольку оно может приводить к неправильной сборке документов.

Для определения, состоящего из нескольких разделов, кроме последовательности и числа повторений разделов, можно задать проверку **ключевых полей**. Для этого в каждом разделе документа определяются ключевые поля, значения которых в разных разделах должны быть одинаковыми. В качестве ключевого поля можно взять, например, фамилию заполняющего документ, которую он должен написать на каждой странице, или некий идентификационный номер.

При обработке пакета программа пытается собрать в документы следующие друг за другом страницы, производя при этом проверку правил сборки. Будет произведена проверка значений ключевых полей. При несовпадении ключевых полей на страницах одного документа, появится сообщение об ошибке. Иногда это происходит из-за того, что страницы документов перепутаны при сканировании, и для выполнения правил сборки документов достаточно изменить порядок следования страниц.

Для того, чтобы добавить страницу в раздел, воспользуйтесь пунктом меню редактора определения документа **Определение документа > Добавить страницу...** или пунктом локального меню изображения **Добавить страницу...** После этого загрузите изображение новой незаполненной страницы и выберите типы объектов, которые должны быть отмечены на странице автоматически.

Для того чтобы добавить новый раздел, воспользуйтесь пунктом меню **Определение документа > Добавить раздел документа...** в окне редактора определения документа. Откроется мастер создания раздела, который поможет вам задать все необходимые параметры. Следуя инструкциям мастера, выполните следующие действия:

- задайте название раздела;
- загрузите изображение незаполненной страницы (отсканируйте или загрузите из файла);
- при необходимости загрузите файл гибкого описания;
- выберите типы объектов, которые будут отмечены автоматически.

Новый раздел и содержащиеся в нем поля отобразятся в окне структуры документа.

Для того чтобы просмотреть получившуюся структуру документа при создании определения, в окне **Структура документа**, расположенном в правой части редактора описания документа, перейдите на закладку **Страницы**. На ней представлена структура документа в режиме эскизов страниц. На этой закладке можно изменить количество вхождений разделов в документ (для этого введите новое значение – число или интервал – в квадрате справа от названия раздела), подключить страницы приложения, изменить порядок страниц в разделе и перенести страницы из одного раздела в другой простым перетаскиванием изображения страницы.

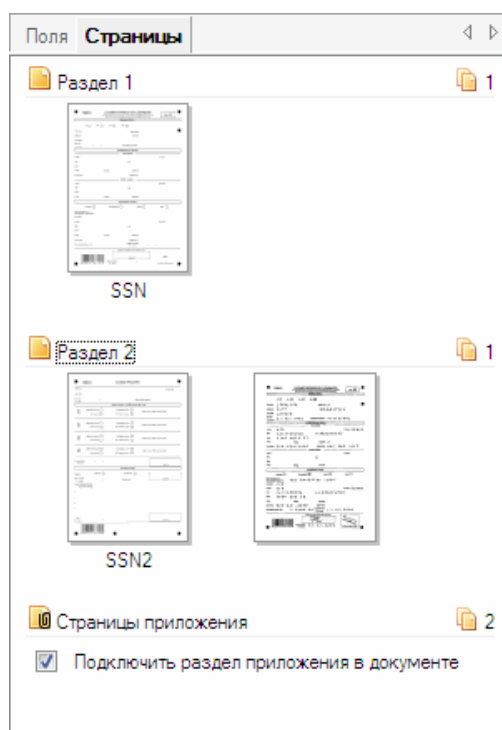


Рисунок 14. Структура документа, режим эскизов страниц

Вы также можете задать последовательность и число повторений разделов, воспользовавшись пунктом меню **Определение документа > Свойства описания документа...** в окне редактора определения документа.

На закладке **Сборка документов** (Рисунок 15) задайте минимальное и максимальное число вхождений каждого раздела в документ (по умолчанию 1).

Если вы хотите проверять значения ключевых полей, отметьте опцию **Проверять совпадение ключевых полей** и выберите ключевое поле в каждом разделе. Ключевое поле выбирается с помощью двойного нажатия левой клавиши мыши.

В некоторых случаях бывает необходимо отключить проверку порядка следования разделов в документе, например, если порядок не важен для сборки документов. Для этого отметьте опцию **Не проверять порядок разделов в документе**. Наличие всех разделов в документе в любом случае будет проверяться, а порядок следования разделов - нет.

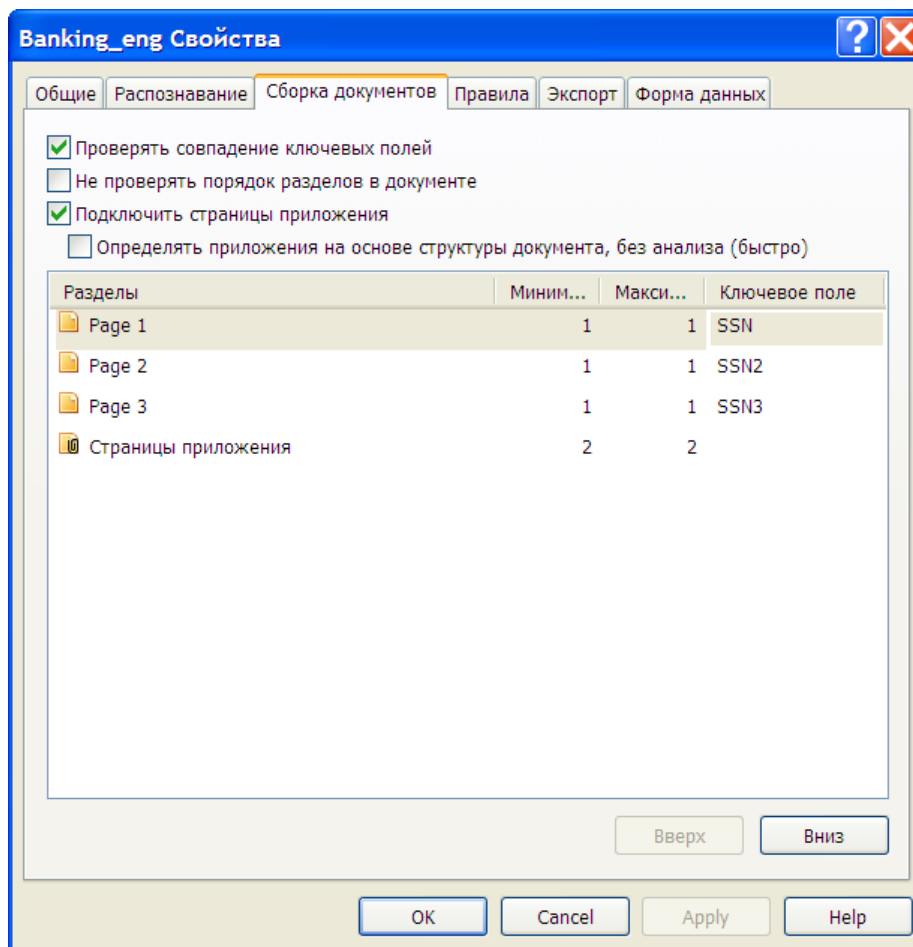


Рисунок 15. Диалог для задания структуры и правил сборки документов

3.1.5. Как создать определение документа с приложениями

ABBYY FlexiCapture 9.0 позволяет создавать описания для документов с приложениями.

Приложения – это дополнительные страницы, которые могут сопровождать каждый документ. С этих страниц не требуется извлекать отдельные поля, поэтому наложения на них определений не требуется, но они учитываются при сборке документов. Например, к заявлению на получение кредита (жесткая форма) прилагается справка с места работы, составленная в свободном стиле. Эта справка может обрабатываться как приложение.

Чтобы создать определение документа с приложением:

- перейдите на закладку **Страницы** в окне **Структура документа** (правая часть редактора определения документа) и включите опцию **Подключить раздел приложения в документе**. Введите количество или диапазон страниц приложения в появившемся квадрате справа от названия раздела. (Рисунок 14)
- или откройте диалог свойств определения документа (**Определение документа > Свойства определения документа...** в окне редактора определения документа) и перейдите на закладку **Сборка документов**. Отметьте опцию **Подключить страницы приложения**. Введите минимальное и максимальное количество страниц в приложении (Рисунок 15).

Для того чтобы сохранить изображение приложения, нужно при настройке экспорта задать опции сохранения изображений (в диалоге **Настройки экспорта** нажмите кнопку **Редактировать**, в открывшемся диалоге выберите **Сохранить изображения документа**). Изображение можно сохранить в каком-либо графическом формате, а также в PDF с возможностью поиска.

3.1.6. **Настройка экспорта данных**

Чтобы задать способ сохранения данных, полученных в ходе обработки документов, необходимо настроить экспорт для каждого определения документа. Возможен экспорт четырех типов: в файл заданного формата, в базу данных, в библиотеку документов MS SharePoint или экспорт, описанный с помощью скрипта. Настройка экспорта осуществляется в окне **Настройки экспорта** (вызывается с помощью пункта **Определение документа > Настройки экспорта** главного меню редактора определения документа). Для добавления нового направления экспорта нажмите кнопку **Добавить...**, для редактирования существующего пути экспорта выберите его в списке и нажмите кнопку **Изменить...**

Наряду с настройками экспорта в определении документа, экспорт можно настроить на уровне типа пакетов и проекта. Экспорт, настроенный на уровне проекта, используется для пакетов с типом «По умолчанию». Данные настройки используются при умолчанию при запуске экспорта оператором с помощью пунктов меню **Файл>Экспортировать в**.

Добавление нового пути экспорта осуществляется с помощью мастера настройки приемника экспорта. На первой странице указываются общие настройки экспорта (тип, обязательность и т. д.), на следующих страницах – настройки для выбранного типа экспорта.

Для одного определения документа можно задать несколько путей экспорта. Путь экспорта может быть обязательным или необязательным. Если при экспорте по обязательному пути возникает сбой, то экспорт завершается с ошибкой, документ не считается экспортированным. Если сбой возникает при экспорте по необязательному пути, ошибки экспорта не возникает, документ будет помечен как экспортированный. Таким образом, желательно, чтобы в каждом определении документа был указан хотя бы один обязательный путь экспорта – в противном случае возникновение сбоев при экспорте останется незамеченным. Признак обязательности пути экспорта устанавливается в списке путей экспорта в окне **Настройки экспорта**.

Изображения страниц также можно экспортировать, как совместно с данными, так и в виде отдельного варианта экспорта. При этом для каждого пути экспорта можно указать, какую информацию на изображении необходимо скрыть. Для экспорта изображений совместно с данными отметьте опцию **Экспортировать изображения документа** на первой странице мастера настройки приемника экспорта.

Примечание. Для корректного экспорта в распределенной версии системы в выгруженном на сервер проекте пути к файлам должны быть указаны в формате UNC (\\server_name\share_name[path]). Если указаны относительные пути, результаты экспорта будут доступны только с помощью программного интерфейса (API). Если указаны абсолютные пути в формате <disk>:\[path], то экспорт будет завершаться ошибкой.

Если указаны относительные пути экспорта, то при выгрузке проекта на сервер можно указать корневой путь в формате UNC. В этом случае относительные пути будут отсчитываться от него.

Направление можно задать для экспорта документов без ошибок, с ошибками или для всех документов. Чтобы указать, какие документы должны экспортироваться, используйте опции **Документы без ошибок** и **Документы с ошибками** в разделе **Экспортировать**.

Экспорт может осуществляться автоматически либо по запросу оператора. В распределенной версии системы для того, чтобы экспорт осуществлялся по запросу оператора, необходимо в настройках проекта создать очередь подтверждения экспорта. В автономной версии системы для того, чтобы экспорт осуществлялся автоматически, необходимо отметить опцию **Разрешить автоматический экспорт** в окне **Настройки экспорта**.

3.1.6.1. Экспорт в файл

Если вы хотите настроить экспорт в файл, выберите значение **Экспорт в файлы** в поле **Тип**. Укажите общие настройки пути экспорта и нажмите кнопку **Далее**.

В поле **Путь** введите путь к папке для хранения экспортированных документов или нажмите кнопку **Обзор...**, чтобы указать путь к папке в дереве каталогов.

Вы можете экспортировать документы одного пакета в один файл или каждый документ в отдельный файл (опция **Экспортировать каждый документ в отдельный файл**). Для каждого пакета можно создавать отдельную папку (опция **Создавать папку для каждого пакета**).

Отметьте опцию **Разрешить перезапись файлов**, если необходимо, чтобы при экспорте новые файлы заменяли старые с тем же именем.

Задайте критерии формирования имен файлов экспорта. Для этого нажмите на кнопку **Задать имена файлов...** и введите нужные критерии в открывшемся диалоге.

Нажмите кнопку **Далее** для перехода к следующей странице.

Укажите тип файла (поддерживаются форматы CSV, DBF, TXT, XLS, XML) и задайте дополнительные свойства экспорта для выбранного типа файла, например, кодировку текста.

Нажмите кнопку **Далее** для перехода к следующей странице.

Если изображения экспортируются вместе с данными, настройте параметры экспорта изображений (аналогично описанному в параграфе «Экспорт изображений»).

На последней странице мастера укажите название пути экспорта.

3.1.6.2. Экспорт в базу данных

Для настройки экспорта в базу данных выберите значение **Экспорт в ODBC-совместимую базу данных** в поле **Тип**. Укажите общие настройки пути экспорта и нажмите кнопку **Далее**.

Введите **Строку соединения** или нажмите на кнопку **Установить...** и задайте параметры в открывшемся диалоге (**Свойства связи с данными**). Далее выберите схему из выпадающего списка.

Чтобы проверить соединение с базой данных, нажмите на кнопку **Тестировать**.

Далее, необходимо определить, в какие таблицы и столбцы таблиц базы данных будут экспортироваться значения полей описания документа. Для этого нажмите кнопку **Соответствие полей...**

В диалоге **Соответствие полей** слева расположены разделы и поля документа, в правой части им нужно поставить в соответствие таблицы и поля базы данных. Если в базе данных уже есть таблицы для экспорта данных, то в диалоге **Соответствие полей** выберите таблицу базы данных для каждого раздела и столбец таблицы базы данных для каждого поля документа. Формат столбца таблицы должен соответствовать значению экспортируемого поля.

Если в базе данных нет таблиц для экспорта документов, можно создать таблицы базы данных автоматически при первом запуске экспорта, для этого надо нажать на кнопку **Создать таблицы автоматически**. При ее нажатии создаются таблицы именно той структуры, которые нужны для экспорта. Полям документа будут поставлены в соответствие столбцы таблиц.

Группы полей, поля с несколькими экземплярами, таблицы документа, разделы экспортируются в отдельные таблицы базы данных. Для связи между главной и подчиненными таблицами используются первичный ключ (Primary key) и внешний ключ (Foreign key). В главной таблице к каждой записи добавляется первичный ключ, а в подчиненной таблице к каждой записи добавляется внешний ключ, в который записывается значение первичного ключа той записи из главной таблицы, к которой относится данная запись. Первичный и внешний ключи добавляются автоматически, где необходимо. Нужно только указать, в каком поле должен быть создан ключ.

Опция **Показывать связанные столбцы** отмечена по умолчанию, если с нее снять отметку, то в списках доступных колонок таблиц в правой части дерева не будут отображаться колонки, в которые уже настроен экспорт.

Нажмите кнопку **Далее** для перехода к следующей странице, где настраивается экспорт изображений.

Вы можете сохранять изображения в базу данных или в виде файлов. Если изображения экспортируются в виде файлов, настройте параметры экспорта изображений (аналогично описанному в параграфе «Экспорт изображений»).

Примечание. Если выбрано сохранение изображений в базу данных, то для них создается столбец с типом BLOB. Если в дальнейшем изменить экспорт изображений, направив их в папку, то при экспорте возникнет ошибка, так как в поле типа BLOB будет экспортироваться текстовый путь к файлу изображения.

На последней странице мастера укажите название пути экспорта.

3.1.6.3. Экспорт в библиотеку документов SharePoint

ABBYY FlexiCapture 9.0 позволяет настроить экспорт документов в библиотеку Microsoft™ SharePoint. В результате экспорта, документы будут сохранены в библиотеке SharePoint, каждому документу будут соответствовать колонки со значениями из определенных полей документа. Эти значения можно использовать для поиска и индексации документов.

Замечания.

1. Для настройки экспорта в SharePoint необходимо иметь права администратора, для осуществления экспорта достаточно прав участника (contributor).
2. Колонки в SharePoint, в которые планируется записывать данные, должны иметь тип *Single line of text* или *Multiple lines of text*.

Если вы хотите настроить экспорт в SharePoint, выберите значение **Экспорт в SharePoint** в поле **Тип**. Укажите общие настройки пути экспорта и нажмите кнопку **Далее**.

Введите адрес сервера (**URL сервера**), где находятся библиотеки SharePoint. С помощью кнопки **Настройки соединения...** можно настроить параметры аутентификации (по умолчанию используются параметры аутентификации для входа в Windows) и настройки Proxy, если это необходимо.

С помощью кнопки **Соединение** можно восстановить соединение с сервером, если его нет.

Выберите библиотеку SharePoint из списка.

Выберите тип содержимого для документов в поле **Тип содержимого** (поддержано начиная с SharePoint 2007). В этом случае экспорт значений можно будет настроить в поля, относящиеся к выбранному типу содержания. Нажмите на кнопку **Соответствие полей...** и произведите необходимые настройки в открывшемся диалоге **Соответствие полей**.

Нажмите кнопку **Далее**.

Настройте параметры экспорта изображений в файлы и нажмите кнопку **Далее**.

На последней странице мастера укажите название пути экспорта.

3.1.6.4. Экспорт изображений

Если вы хотите настроить экспорт изображений, выберите значение **Экспорт файлов изображений** в поле **Тип**. Укажите общие настройки пути экспорта и нажмите кнопку **Далее**.

В поле **Путь** введите путь к папке для хранения экспортированных изображений или нажмите кнопку **Обзор...**, чтобы указать путь к папке в дереве каталогов.

Вы можете экспортировать документы одного пакета в один файл (опция **Создавать папку для каждого пакета**). Отметьте опцию **Разрешить перезапись файлов**, если вы допускаете перезапись файла экспорта.

Задайте критерии формирования имен файлов экспорта. Для этого нажмите на кнопку **Задать имена файлов...** и введите нужные критерии в открывшемся диалоге.

Если необходимо скрывать на экспортируемых изображениях какую-нибудь информацию, отметьте опцию **Удалять конфиденциальную информацию**. Далее нажмите кнопку **Выбрать поля для удаления...** и в открывшемся диалоговом окне укажите список полей, которые необходимо скрывать.

Нажмите кнопку **Далее** для перехода к следующей странице.

Выберите формат хранения изображений в поле **Тип файла**. Укажите качество и цветность сохраняемых изображений.

Если вам нужно изменить разрешение исходных изображений, например, чтобы уменьшить объем хранимых данных, отметьте опцию **Изменить разрешение на** и введите разрешение.

Если вы выберете формат PDF и отметите опцию **Создать PDF с возможностью поиска**, то будет сохранено не просто изображение документа, а будет произведено его полнотекстовое распознавание, а распознанный текст сохранится в выбранном формате. В этом случае вы сможете задать язык распознавания: оставить как в описании документа или выбрать из списка один или несколько языков (кнопка **Выбрать**).

Нажмите кнопку **Далее**.

На последней странице мастера укажите название пути экспорта.

3.1.6.5. Экспорт, описанный с помощью скрипта

Используя скриптовые процедуры, вы можете настроить экспорт более гибко, чем это позволяют средства, предоставляемые интерфейсом пользователя.

Если вы хотите настроить экспорт с помощью скриптовой процедуры, выберите значение **Пользовательский экспорт (скрипт)** в поле **Тип**. Укажите общие настройки пути экспорта и нажмите кнопку **Далее**.

Выберите язык скрипта (**JScript®** или **VBScript**) и введите текст скрипта в окне редактора, которое появляется при нажатии кнопки **Редактировать...** (Подробное описание скриптового языка с примерами использования вы найдете в справочной системе.)

3.1.7. *Настройка отображения распознанных данных*

После того, как данные будут распознаны, оператор увидит их в окне документа. По умолчанию данные будут расположены в виде списка, подписи к ним будут соответствовать названиям полей определения документа. Однако можно изменить представление данных, расположив их удобным для пользователя способом или изменив заголовки элементов. Можно также добавить любой пояснительный текст. Представление данных можно изменить в нижней правой части окна редактора определений документов.

Поля можно перемещать, менять их названия или другие свойства. Чтобы добавить текст, выберите пункт локального меню **Вставить текстовую метку**.

Шрифт и размер текста для отображения названий полей и распознанных значений данных можно задать на закладке **Форма данных** диалога **Свойства определения документа (Определение документа > Свойства определения документа...)**.

3.1.8. *Тестирование определения документа*

Прежде чем приступить к массовому вводу документов, необходимо протестировать определения документов.

Тестировать описание можно прямо в редакторе описания документа, на том изображении, которое используется для создания описания. Чтобы начать тестирование, выберите пункт меню **Тестирование>Начать тестирование**.

Дальнейшее тестирование производится на документах, добавляемых в тестовые пакеты, т.к. эти пакеты работают с неопубликованной версией определения. Эти пакеты доступны из меню **Вид>Тестовые пакеты** главного окна программы. Также к ним можно перейти из окна редактора описания документа, выбрав пункт меню **Пакет>Тестовые пакеты**.

Скриптовые правила можно тестировать непосредственно из окна редактора кода скрипта. Для этого выберите в меню пункт **Сервис>Начать тестирование**. Данную возможность можно использовать, чтобы проверить поведение скрипта при различных значениях полей.

Если в процессе тестирования были обнаружены ошибки в настройках правил, некорректно заданные значения свойств полей и т.д., то для их исправления отредактируйте описание. После устранения всех найденных проблем можно опубликовать описание документа и приступить к массовому вводу документов.

3.1.9. Редактирование и публикация определения документа

После создания определения документа его нужно протестировать на нескольких изображениях. Получив удовлетворительный результат, определение нужно опубликовать, чтобы оно стало доступно для распознавания рабочих пакетов.

Чтобы опубликовать определение документа, нажмите на кнопку **Опубликовать** в диалоге **Определения документов** (меню **Проект > Определения документов...** главного окна программы).

Чтобы вернуться к редактированию определения документа, выберите его в диалоге **Определения документов** и нажмите на кнопку **Правка...** Определение будет недоступно для редактирования другими пользователями, а распознавание будет производиться с использованием последней опубликованной версии, так что редактирование определения не помешает работе операторов. После редактирования опубликуйте определение документа. После опубликования пользователям станет доступна новая версия. Если вы не хотите публиковать полученную версию, вы также можете отказаться от сделанных изменений и вернуться к последней опубликованной версии. Для этого в диалоге **Определения документов** нужно выделить определение и нажать на кнопку **Отменить правку**.

После публикации новой версии определения документы будут распознаваться с ее помощью. Для отдельных распознанных ранее документов можно изменить используемую версию определения без перераспознавания (если изменения между версиями не касались положения полей и настроек распознавания, а заключались, например, в изменении настроек экспорта). Для этого используйте пункт главного меню **Сервис > Обновить до последней версии**.

3.2. Настройка системы для ввода гибких документов

Более сложной задачей является настройка процесса обработки гибких документов. Для счетов, платежных поручений и других аналогичных документов невозможно создать определение с жестким расположением полей, потому что местоположение, размеры и количество полей на разных экземплярах могут не совпадать, а значит, жесткое описание не может быть наложено на все документы. На таких документах поля ищутся после предварительного распознавания всей страницы целиком, а описание создается на основе ключевых слов и взаимного расположения полей. Для разработки таких описаний предназначен специальное приложение АБВУ FlexiLayout Studio 9.0. Подробное руководство создания гибкого описания вы найдете в справке для АБВУ FlexiLayout Studio 9.0.

Созданное гибкое описание экспортируется в .afl файл, а затем подключается в АБВУ FlexiCapture 9.0. Гибкое описание подключается в качестве одного из разделов определения

документа. Таким образом, определение может содержать одновременно и жесткие, и гибкие разделы. Гибкое описание может быть многостраничным.

Гибкое описание можно подключить на этапе создания определения документа. Для этого на втором шаге Мастера создания определений документов добавьте изображение документа, а затем отметьте опцию **Загрузить гибкое описание** и введите путь к файлу с расширением .afl, содержащему гибкое описание.

Вы можете подключить гибкое описание и из окна редактора определения документа, в диалоге свойств раздела определения документа. Диалог свойств раздела открывается при выборе пункта **Свойства...** локального меню раздела. Перейдите на закладку **Гибкое описание** и нажмите **Загрузить...** Выберите нужный .afl файл, содержащий гибкое описание.

После подключения гибкого описания все поля и их геометрическая разметка полей уже будут присутствовать на изображении.

Набор полей в определении документа должен строго соответствовать набору полей в гибком описании. Не следует удалять и добавлять поля (за исключением полей, не имеющих области на изображении), набор распознаваемых полей можно изменять, только изменив само гибкое описание.

Теперь нужно задать свойства полей точно так же, как и для жесткого описания. Определите свойства распознавания, верификации, экспорта, задайте необходимые правила.

3.2.1. Классификаторы

Классификатор – специальный проект ABBYY FlexiLayout Studio, предназначенный для предварительной идентификации отдельных страниц и отнесения их к определенному типу (например, по используемому гибкому описанию / варианту гибкого описания). После разработки классификатор используется в ABBYY FlexiCapture для классификации документов на уровне типа пакета (для типа пакета «По умолчанию» классификатор подключается на уровне проекта).

Классификаторы могут использоваться в следующих целях:

- Для ускорения выбора определения документа для наложения. Использование классификатора позволяет не перебирать все определения документа для поиска наиболее подходящего, а выбирать нужное определение на основании классификационных признаков. При этом порядок перебора определяется пользователем с помощью выбора классификационных признаков и построения дерева классификатора. Ускорение наиболее заметно для крупных проектов, содержащих большое количество определений документов.
- Для распределения классифицирующих признаков по дереву классификатора. На верхнем уровне можно отбирать документы, поступившие от определенной компании, по наличию поля с ее наименованием. На следующем уровне классификации отобранные документы можно различать по типам с помощью заголовков или, например, количества разделителей на странице.
- Для управления наложением гибких описаний сторонней разработки. Возможна ситуация, когда описания, полученные от стороннего разработчика, накладываются неправильно (не на те документы, на которые требуется). В данном случае классификатор можно использовать для управления наложением определений документов, не обращаясь к разработчику описаний.
- Для ускорения работы проектов, содержащих определения документов на разных языках. При использовании классификатора предраспознавание для классифицируемых документов

будет осуществляться не по количеству языков всех определений в проекте, а два раза: при классификации и при наложении выбранного определения документа.

Подробная информация о классификаторах приведена в справочной системе ABBYY FlexiLayout Studio.

Классификаторы могут подключаться на уровне типа пакета или проекта в целом (во втором случае классификатор действует для пакетов с предустановленным типом «По умолчанию»).

Для подключения классификатора выполните следующие действия:

- Выберите пункт **Проект > Свойства проекта...** меню главного окна для загрузки классификатора на уровне проекта либо откройте свойства типа пакета (пункт **Проект > Типы пакетов...**, в открывшемся окне **Типы пакетов** - кнопка **Правка...**) для загрузки классификатора на уровне типа пакета;
- в открывшемся диалоге свойств перейдите на закладку **Распознавание** и нажмите кнопку **Загрузить...**;
- Выберите *.cfl файл, содержащий классификатор.

3.3. Особенности ввода неструктурированных документов

Если вам нужно обрабатывать неструктурированные документы, содержащие информацию, изложенную в свободном стиле, например, контракты, письма, приказы, приложения, то и в этом случае поможет ABBYY FlexiCapture 9.0. Неструктурированные документы с текстом или изображениями, разделенные пустыми страницами или страницами со штрих-кодом, обрабатываются и экспортируются в PDF файлы с возможностью поиска или в файлы графических форматов.

Обработка таких документов обычно состоит в том, чтобы перевести документы в электронный вид и организовать их поиск по значениям ключевых полей.

Обычно, поиск ключевых полей (например, номер контракта) на таких документах производится с помощью гибкого описания, созданного в приложении ABBYY FlexiLayout Studio.

Если автоматический поиск ключевых полей невозможен, то их значения могут вводиться вручную оператором. Для этого создайте описание документа с одним полем (или с несколькими полями, если нужно), и в свойствах распознавания этого поля (полей) отметьте опцию **Не распознавать значение поля, оно будет введено вручную**. В этом случае при запуске процесса верификации оператору будет предложено ввести значение ключевого поля (или полей) вручную.

Для хранения документов нужно настроить экспорт: значения ключевых полей вы можете экспортировать в файл или базу данных, а изображения документов сохранить в удобном для вас формате. Для экспорта изображений на закладке **Изображения** диалога **Настройки экспорта** отметьте опцию **Сохранять изображения документов** и задайте параметры сохранения изображений. Вы можете сохранять изображения документов в графических файлах или в PDF с возможностью поиска.

Обратите внимание на сборку страниц в документы: для неструктурированных документов бывает сложно определить, к какому документу относится та или иная страница. Для автоматизации сборки неструктурированных документов рекомендуется разделять документы чистыми страницами или страницами со штрих-кодами. Тогда при добавлении изображений в пакет (путем сканирования, добавления из файлов или при создании профиля импорта) нужно отметить опцию

Для изображений, разделенных и выбрать из выпадающего списка значение **пустыми страницами** или **страницами со штрих-кодом** в зависимости, от того, какие страницы используются в качестве разделителя. Сборка страниц в документы будет произведена автоматически: страницы будут добавляться в текущий документ, пока не встретится страница-разделитель.

4. Настройка свойств проекта

Настройка свойств проекта осуществляется в окне **Свойства**. Чтобы открыть данное окно, выберите пункт **Проект > Свойства проекта...**

Окно содержит шесть закладок. Закладка **Общие** содержит описание общих свойств проекта – название, описание и т.д. Остальные закладки аналогичны одноименным закладкам окна **Свойства типа пакета** (см. раздел «Настройка типов пакетов») и описывают настройки пакетов с предустановленным типом «По умолчанию».

На закладке **Общие** можно настроить следующие свойства проекта:

- **Название.**
- **Описание.**
- **Корневой путь для экспорта** — папка, от которой будут определяться относительные пути для экспорта в файлы после загрузки проекта на сервер. Путь указывается в формате UNC.

Также на закладке **Общие** можно просмотреть название файла проекта, включая путь в дереве папок (для проекта на сервере приложений – его GUID).

Остальные закладки аналогичны настройке типов пакетов и используются для настройки типа пакета «По умолчанию».

При настройке свойств проекта можно указать список и порядок используемых определений документов. В отличие от типов пакетов, данная операция осуществляется в окне **Определения документов** (пункт **Проект > Определения документов...** главного меню приложения).

Гибкие определения документов накладываются в том же порядке, в котором они перечислены в данном окне (сверху вниз). Для изменения положения определения документа в списке используйте кнопки **Вверх** и **Вниз**.

Обычно при распознавании последовательно накладываются все подключенные гибкие описания, после чего сравнивается качество их наложения и выбирается наилучшее. Для ускорения наложения определений можно выбрать опцию **Использовать первое подошедшее гибкое описание**. В этом случае при наложении какого-либо гибкого описания дальнейшего перебора осуществляться не будет. Данная опция полезна при большом количестве непохожих гибких описаний, поскольку снижает общее время распознавания.

При необходимости можно временно отключить какое-либо определение документов. Для этого снимите опцию **Включен** в списке определений. Для подключения определения включите данную опцию.

5. Настройка типов пакетов

В ABBYY FlexiCapture 9.0 предусмотрен пакетный режим обработки документов: перед началом обработки создается пакет, в него добавляются изображения документов и далее они могут

обрабатываться в этом пакете. Пакеты создаются автоматически или вручную при импорте изображений (сканировании или добавлении уже отсканированных изображений документов). Настройки, с которыми будут обрабатываться пакеты документов, задаются в типе пакета. Можно создать несколько типов пакетов с наиболее часто используемыми наборами настроек обработки. Тогда, чтобы указать с какими настройками должен обрабатываться пакет, оператору необходимо просто задать тип пакета.

Тип пакета определяет следующие параметры обработки документов:

- способ предварительной обработки изображений;
- список и порядок наложения определений документа также используемый классификатор;
- проверка целостности пакета;
- настройки «по умолчанию» для экспорта в ручном режиме;
- настройки экспорта нераспознанных документов;
- процесс обработки пакета.

Также тип определяет список регистрационных параметров пакета - текстовой информации, характеризующей конкретный пакет.

Типы пакетов создаются в рамках определенного проекта. Для просмотра списка типов пакета используется окно **Типы пакетов** (пункт **Проект > Типы пакетов** главного меню).

Для создания нового типа пакета нажмите кнопку **Новый...** Далее укажите параметры пакета с помощью мастера создания нового типа пакетов.

Для редактирования пакета выберите его в списке и нажмите кнопку **Правка...** Будет открыто диалоговое окно **Свойства типа пакета**.

Настройки типа пакета объединены в шесть групп. Каждой группе соответствует своя страница мастера создания нового типа пакета и закладка в окне **Свойства типа пакета**.

5.1. Общие свойства

К общим относятся следующие свойства:

- **Название** типа пакетов.
- **Описание** типа пакетов.
- **Регистрационные параметры.** Регистрационные параметры – текстовые свойства, позволяющие идентифицировать конкретный пакет. В качестве таких свойств можно использовать, например, дату сканирования, фамилию оператора сканирования и т.д. Регистрационные параметры указываются при создании пакета выбранного типа. Регистрационный параметр может быть обязательным и необязательным для указания.

5.2. Обработка изображений

Для типа пакета можно настроить следующие параметры создания новых документов и обработки изображений:

- **Создание нового документа.** Новый документ может создаваться автоматически при наложении определения документа, для каждого файла или при появлении страницы-разделителя. В качестве страницы разделителя может использоваться пустая страница или страница со штрих-кодом определенного типа или с указанным значением. Если страницу-

разделитель не требуется обрабатывать, ее можно удалять, выбрав опцию **Удалять страницу-разделитель**.

- **Обработка изображения.** Можно указать следующие параметры обработки изображений:
 - **Очищать изображения,**
 - **Конвертировать цветные и серые изображения в черно-белые,**
 - **Разворачивать изображения** (на 90° по и против часовой стрелки либо на 180°),
 - **Выравнивать изображения по разделителям,**
 - **Выравнивать изображения по черным квадратам.**

Данные настройки будут использоваться по умолчанию при ручном добавлении операторами документов в пакет этого типа без использования профилей импорта.

5.3. Распознавание

Для типа пакета можно настроить следующие параметры распознавания:

- **Используемые определения документов.** Можно использовать все определения документов, входящие в проект, а можно указать список определений непосредственно для типа пакета. При использовании списка определений документа для типа пакета можно указать порядок их наложения: определения накладываются в том порядке, в котором они перечислены в списке (сверху вниз). Для изменения положения определения в списке выберите его и используйте кнопки **Выше** и **Ниже**.
- **Выбор гибкого описания.** Обычно при распознавании последовательно накладываются все подключенные гибкие описания, после чего сравнивается качество их наложения и выбирается наилучшее. Для ускорения наложения определений можно выбрать опцию **Использовать первое подошедшее гибкое описание**. В этом случае при наложении какого-либо гибкого описания дальнейшего перебора осуществляться не будет.
- **Дерево классификатора.** Для типа пакета, как и для проекта в целом, можно подключить классификатор для выбора гибких описаний. Для выбора классификатора нажмите кнопку **Загрузить...** и в открывшемся диалоговом окне выберите файл классификатора. Для отключения классификатора нажмите кнопку **Отключить**.
- **Удалять пустые страницы при распознавании** – выберите опцию, если пустые страницы необходимо удалять.

5.4. Проверка целостности

Если необходимо проверять целостность пакетов данного типа, выберите опцию **Выполнять проверку целостности пакета перед экспортом**. Для редактирования скрипта проверки целостности нажмите кнопку **Правка...** При этом будет автоматически отмечена и недоступна для изменения опция **Экспортировать пакеты целиком** на закладке **Процесс обработки** (в распределенной версии).

Если текст скрипта не указан, то проверка целостности выполняться не будет вне зависимости от состояния опции.

5.5. Экспорт

Параметры экспорта указываются для определения документа. Для типа пакета можно указать параметры экспорта по умолчанию в ручном режиме. Текущие настройки экспорта отображаются в

одноименном блоке. Для изменения настроек экспорта нажмите кнопку **Изменить настройки по умолчанию...** и укажите параметры экспорта в диалоговом окне (см. параграф «Настройка экспорта данных»).

При необходимости после завершения экспорта можно автоматически запускать скрипт. Для этого отметьте опцию **Выполнять скрипт после экспорта**. Для перехода к редактированию скрипта нажмите кнопку **Правка...** Подробно язык скриптов описан в справочной системе.

Для экспорта изображений нераспознанных документов (на которые не наложилось ни одного определения) отметьте опцию **Экспортировать нераспознанные документы**. Для изменения параметров экспорта изображений нажмите кнопку **Изменить...** и укажите новые параметры (см. пункт «Экспорт изображений»).

5.6. Процесс обработки

5.6.1. Автономная версия

В автономной версии системы можно настроить следующие параметры процесса обработки:

- Автоматический экспорт пакетов при малом количестве неуверенно распознанных символов. Для того, чтобы пакеты экспортировались автоматически, отметьте опцию **Экспортировать пакеты автоматически при отсутствии ошибок и предупреждений и количестве сомнительных символов менее или равно, чем** и укажите максимальный процент неуверенно распознанных символов, при котором допускается автоматический экспорт.
- Удаление документов после экспорта. Для этого отметьте опцию **Удалять обработанные документы после экспорта**.

5.6.2. Распределенная версия

В распределенной версии системы можно настроить следующие параметры обработки:

- **Режим обработки:** автоматический или с верификацией. В автоматическом режиме система работает без участия человека, все действия осуществляются автоматически: документы импортируются, распознаются и экспортируются, при возникновении ошибок документы попадают в очередь Исключений. В режиме с верификацией человек осуществляет проверку и исправление неуверенно распознанных символов и ошибок, а также при необходимости запускает экспорт.
- **Верифицировать пакеты целиком.** Задания на верификацию формируются из определенного (10 или менее) документов из одного пакета. При выборе данной опции задания на верификацию будут включать в себя пакет целиком.
- **Пропускать верификацию и верификацию данных, если неуверенно распознанных символов в документе не более, чем** — при выборе данной опции документы с хорошим качеством распознавания не включаются в задания на верификацию. Документы будут пропускать верификацию, если в них нет символов для верификации и процент неуверенно распознанных символов ниже указанного значения.
- **Экспортировать пакеты целиком.** При выборе данной опции пакеты документов будут экспортироваться только целиком, экспорт отдельных документов осуществляться не будет.
- **Удалять документы после экспорта** — отметьте данную опцию для удаления экспортированных документов.

- **Очереди.** Подключение дополнительных очередей, в которых будут формироваться задания операторам.

6. Настройка импорта изображений

В распределенной версии системы импорт изображений осуществляется с помощью специального приложения — Станции сканирования. Однако у операторов верификации и старших операторов верификации также есть возможность добавить изображения в пакет. Кроме того, импорт изображений может осуществляться автоматически из горячих папок.

Профили импорта позволяют указать параметры импорта и обработки изображений. В распределенной версии профили импорта, как правило, используются для обработки горячих папок, так как добавление изображений операторами верификации используется для исправления ошибок сборки, а не для поточной обработки документов. В автономной версии оператор также может использовать профили для быстрого импорта изображений с помощью одной кнопки.

Чтобы добавить профиль импорта, выберите пункт меню **Проект > Профили импорта изображений...** главного окна программы. В открывшемся диалоге вы можете создавать новые профили импорта, редактировать, удалять и копировать существующие профили (Рисунок 16).

Создать новый профиль импорта вам поможет мастер создания профиля, который запустится при нажатии кнопки **Новый...**

На первом шаге мастера выберите источник импорта. Источником импорта может служить сканер или горячая папка (то есть папка, в которой будет производиться поиск новых изображений). Если вы хотите добавлять изображения из горячей папки, убедитесь, что у вас есть права на чтение и запись в данную папку. Если источником импорта будет сканер, убедитесь, что он правильно подключен.

Создание профиля импорта с использованием сканера

1. На первом шаге мастера выберите **Сканер** в качестве источника импорта.
2. На следующем шаге мастера вы можете задать опции сканирования и обработки поступающих изображений. В выпадающем списке **Стиль диалога настроек** можно выбрать способ взаимодействия программы со сканером: **Настройки сканирования FlexiCapture** – для настройки опций сканирования будет использоваться диалог программы ABBYY FlexiCapture или **Системные настройки сканирования** – для настройки опций сканирования будет использоваться диалог соответствующего драйвера сканера.
3. На третьем шаге вы можете задать опции для сборки страниц в документы и опции предварительной обработки изображений. Например, можно указать, что изображения нужно очищать от мусора (это нужно, если изображения не очень качественные и на нем присутствует много лишних точек), конвертировать в черно-белые, разворачивать в указанном направлении или отключить выравнивание.
4. На последнем шаге мастера вы можете изменить имя, присвоенное профилю импорта по умолчанию, и ввести комментарий.

Создание профиля импорта с использованием горячей папки

1. На первом шаге мастера выберите **Горячая папка** в качестве источника импорта и задайте путь к горячей папке. Путь предпочтительно указывать в формате UNC.

2. На втором шаге задайте опции загрузки изображений. Вы можете отметить опцию **Проверять Горячую папку каждые** и ввести промежуток времени, через который будет проверяться наличие новых изображений в горячей папке. В области **Настройки пакета** выберите одну из опций, чтобы определить, в какой пакет будут добавлены новые изображения. В выпадающем списке **Количество файлов** выберите одно из значений, чтобы определить, сколько файлов будет добавлено в указанный или вновь созданный пакет.
3. На следующем шаге, как и при создании профиля импорта с использованием сканера, вы можете задать опции разделения страниц на документы и опции предварительной обработки изображений.
4. Далее, мастер создания профиля предложит вам задать опции очистки горячей папки после импорта: успешно импортированные изображения, как и изображения, в процессе обработки которых произошла ошибка, можно оставлять на месте или перемещать в другую папку.
5. На последнем шаге мастера вы можете изменить имя, присвоенное профилю импорта по умолчанию, и ввести комментарий.

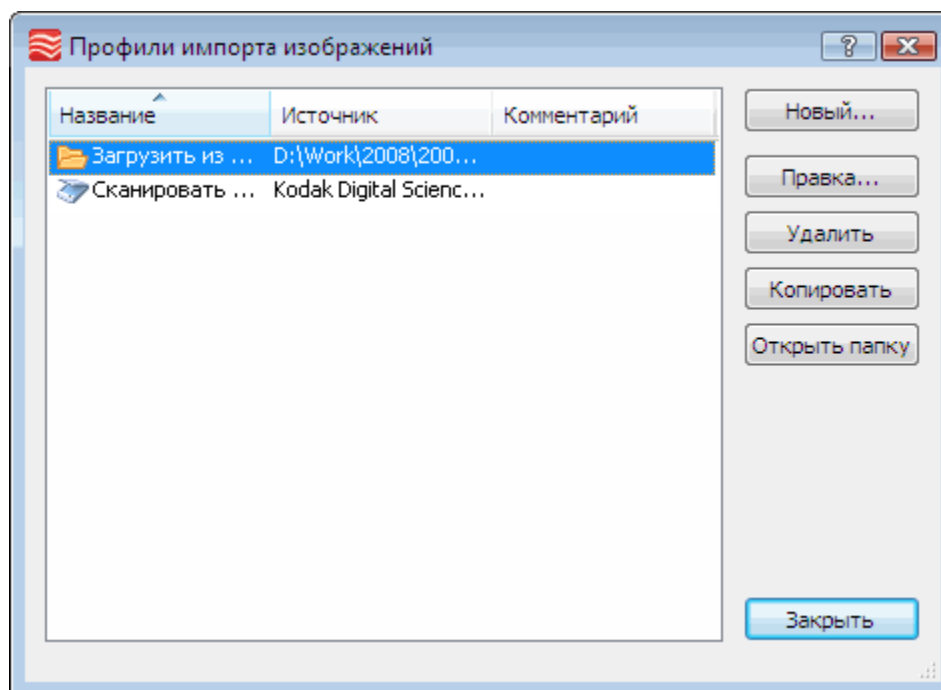


Рисунок 16. Диалог создания и редактирования профилей импорта

Запуск автоматической проверки горячих папок осуществляется следующим образом:

- В автономной версии – путем установки для профиля импорта изображений опции **Автоматическая проверка** в окне **Профили импорта изображений**.
- В распределенной версии – с помощью приложения **Монитор сервера обработки**, в разделе **Горячие папки** для каждого проекта отдельно.

7. Загрузка проекта на сервер

В автономной версии системы после завершения настройки проекта операторы могут сразу начинать с ним работать. В распределенной версии созданный проект необходимо загрузить на сервер приложений. Для этого выберите в меню пункт **Файл > Загрузить проект на сервер...** и в

открывшемся диалоговом окне укажите адрес сервера приложений в формате `http://<servername>`. Загрузить проект на сервер может только администратор сервера приложений.

После загрузки проекта на сервер необходимо присвоить операторам в рамках проекта нужные роли с помощью **Консоли администрирования и мониторинга**. При использовании в проекте горячих папок их необходимо подключить с помощью **Монитора сервера обработки**. Более подробно администрирование системы описано в руководстве администратора.

8. Горячие кнопки

8.1. Главное окно приложения

| | |
|---|--------------|
| Новый проект | Ctrl+Shift+N |
| Открыть проект | Ctrl+Shift+O |
| Сохранить | Ctrl+S |
| Новый пакет | Ctrl+N |
| Загрузить изображения | Ctrl+O |
| Сканировать изображения | Ctrl+K |
| Импортировать изображения | Ctrl+I |
| Экспортировать | Ctrl+U |
| Экспортировать данные в файлы | Alt+Shift+S |
| Экспортировать в базу данных | Alt+Shift+D |
| Отменить | Ctrl+Z |
| Повторить | Ctrl+Y |
| Вырезать | Ctrl+X |
| Копировать | Ctrl+C |
| Вставить | Ctrl+V |
| Удалить | Del |
| Выделить все | Ctrl+A |
| Найти | Ctrl+F |
| Найти следующий | F3 |
| Перейти к следующему документу | Ctrl+D |
| Перейти к предыдущему документу | Ctrl+Shift+D |
| Очистить изображение | Ctrl+Alt+K |
| Инвертировать изображение | Ctrl+Alt+V |
| Повернуть изображение на 90° по часовой стрелке | Ctrl+W |
| Повернуть изображение на 90° | Ctrl+Shift+W |

| | |
|----------------------------------|-----------------|
| против часовой стрелки | |
| Тестовые пакеты | Ctrl+Shift+B |
| Пакеты | Ctrl+B |
| Детали | Alt+1 |
| Эскизы страниц | Alt+2 |
| Полноэкранный режим | F11 |
| Обновить | F5 |
| Свойства | Alt+Enter |
| Редактор документов: Изображение | Ctrl+1 |
| Редактор документов: Данные | Ctrl+2 |
| Увеличить масштаб изображения | Ctrl +Num+ |
| Уменьшить масштаб изображения | Ctrl+Num- |
| Увеличить размер эскизов | Ctrl+Shift+Num+ |
| Уменьшить размер эскизов | Ctrl+Shift+Num- |
| Проанализировать | Ctrl+E |
| Наложить определение документа | Alt+Shift+E |
| Распознать | Ctrl+R |
| Начать верификацию | F7 |
| Следующий объект верификации | F4 |
| Предыдущий объект верификации | Shift+F4 |
| Следующая ошибка сборки | F9 |
| Предыдущая ошибка сборки | Shift+F9 |
| Следующий неуверенный символ | F8 |
| Предыдущий неуверенный символ | Shift+F8 |
| Следующая ошибка правил | F6 |
| Предыдущая ошибка правил | Shift+F6 |
| Определения документов | Ctrl+T |
| Типы пакетов | Ctrl+Shift+T |
| Профили импорта изображений | Ctrl+Shift+I |
| Обновить до последней версии | Alt+Shift+U |
| Повторный анализ | Ctrl+Alt+E |
| Повторное распознавание | Ctrl+Alt+R |
| Перепроверить правила | Ctrl+F6 |
| Проверить целостность пакета | Ctrl+Shift+E |

| | |
|---------------|--------|
| Вызов справки | F1 |
| Закрывать | Alt+F4 |

8.2. Окно редактора определений документов

| | |
|---|--------------|
| Добавить страницу | Ctrl+P |
| Добавить раздел документа | Ctrl+Alt+S |
| Свойства определения документа | Ctrl+T |
| Сохранить | Ctrl+S |
| Отменить | Ctrl+Z |
| Повторить | Ctrl+Y |
| Вырезать | Ctrl+X |
| Копировать | Ctrl+C |
| Вставить | Ctrl+V |
| Удалить | Del |
| Удалить область | Shift+Del |
| Выделить все | Ctrl+A |
| Выделить по типу | Ctrl+Shift+A |
| Сгруппировать | Ctrl+G |
| Разгруппировать | Ctrl+Shift+G |
| Копировать текст с изображения | Ctrl+Alt+C |
| Создать поле: Текст | Alt+Shift+T |
| Создать поле: Метка | Alt+Shift+C |
| Создать поле: Группа меток | Alt+Shift+M |
| Создать поле: Штрих-код | Alt+Shift+B |
| Создать поле: Картинка | Alt+Shift+P |
| Создать поле: Таблица | Alt+Shift+L |
| Создать поле: Группа | Alt+Shift+G |
| Очистить изображение | Ctrl+Alt+K |
| Инвертировать изображение | Ctrl+Alt+V |
| Повернуть изображение на 90° по часовой стрелке | Ctrl+W |
| Повернуть изображение на 90° против часовой стрелки | Ctrl+Shift+W |
| Форма данных | Alt+F2 |

| | |
|---------------------------------|-----------|
| Свойства | Alt+Enter |
| Статические элементы | Ctrl+1 |
| Области полей | Ctrl+2 |
| Показать структуру документа | Alt+F1 |
| Поля | Alt+1 |
| Страницы | Alt+2 |
| Увеличить масштаб изображения | Ctrl+Num+ |
| Уменьшить масштаб изображения | Ctrl+Num- |
| Начать тестирование | Ctrl+R |
| Найти объекты на изображении | Ctrl+D |
| Проверить определение документа | F9 |
| Наложить раздел | Ctrl+E |
| Вызов справки | F1 |
| Закреть | Alt+F4 |

8.3. *Окно групповой верификации*

| | |
|-------------------------------|------------|
| Подтвердить все | Enter |
| Отложить все | Ctrl+Enter |
| Изменить статус | Space |
| Следующая страница | Page Down |
| Предыдущая страница | Page Up |
| Отменить | Ctrl+Z |
| Повторить | Ctrl+Y |
| Выделить все | Ctrl+A |
| Показать изображение символа | F2 |
| Полноэкранный режим | F11 |
| Показывать изображение поля | Ctrl+I |
| Изображение поля: Сверху | Alt+1 |
| Изображение поля: Снизу | Alt+2 |
| Увеличить масштаб изображения | Ctrl+Num+ |
| Уменьшить масштаб изображения | Ctrl+Num- |
| Вызов справки | F1 |
| Закреть | Alt+F4 |

8.4. *Окно контекстной верификации*

| | |
|-------------------------------|-------------|
| Подтвердить поле | Enter |
| Отложить поле | Ctrl+Enter |
| Следующий неуверенный символ | F4 |
| Предыдущий неуверенный символ | Shift+F4 |
| Следующее поле | Page Down |
| Предыдущее поле | Page Up |
| Отменить | Ctrl+Z |
| Повторить | Ctrl+Y |
| Вырезать | Ctrl+X |
| Копировать | Ctrl+C |
| Вставить | Ctrl+V |
| Удалить | Del |
| Удалить все | Alt+Del |
| Выделить все | Ctrl+A |
| Вставить перенос строки | Shift+Enter |
| Объединить символы | Ctrl+M |
| Аналогичные поля | Alt+F3 |
| Показать изображение символа | F2 |
| Полноэкранный режим | F11 |
| Показывать распознанный текст | Alt+F1 |
| Нарезка изображений символов | Alt+F2 |
| Показывать изображение поля | Ctrl+I |
| Изображение поля: Сверху | Alt+1 |
| Изображение поля: Снизу | Alt+2 |
| Увеличить масштаб изображения | Ctrl+Num+ |
| Уменьшить масштаб изображения | Ctrl+Num– |
| Вызов справки | F1 |
| Заккрыть | Alt+F4 |