

**ABBYY FlexiCapture 8.0**  
**Professional**  
**Руководство пользователя**

Уважаемый пользователь!

Данное руководство поможет вам ознакомиться с ABBYY FlexiCapture 8.0 Professional и понять принцип работы с программой. Прочтите это руководство до начала работы.

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 8.0 Professional > Справки.
- Руководство для системного администратора ABBYY FlexiCapture 8.0 Professional. Старт > Программы > ABBYY FlexiCapture 8.0 Professional > Руководства > Руководство для системного администратора.
- Руководство для операторов. Старт > Программы > ABBYY FlexiCapture 8.0 Professional > Руководства > Руководство оператора.
- Руководство по созданию машиночитаемых форм. Старт > Программы > ABBYY FlexiCapture 8.0 Professional > Руководства > Создание машиночитаемых форм.
- Контекстная справка по приложениям FlexiLayout Studio 8.0 и FormDesigner 8.0 вызывается из меню соответствующего приложения или с помощью клавиши F1. Также доступна из меню Старт > Программы > ABBYY FlexiCapture 8.0 Professional > Справки.
- Руководство ABBYY FlexiLayout Studio «Обучение на примерах». Старт > Программы > ABBYY FlexiCapture 8.0 Professional > Руководства > Обучение на примерах.

Чтобы сделать изучение продукта наиболее эффективным, мы предлагаем вам параллельно с чтением руководства настроить проект для обработки конкретного вида документов. В качестве примера мы предлагаем вам для заполнения специально разработанную *Анкету пользователя*. Изображение анкеты находится в папке <имя диска>:\Documents and Settings\All Users\Application Data\ABBYY\FlexiCapture\8.0\Samples\FormDesigner-Russian\Questionnaire (для ОС Microsoft Windows Vista <имя диска>:\Users\Public\ABBYY\FlexiCapture8.0\Samples\FormDesigner-Russian\Questionnaire). По мере чтения руководства вы будете встречать в тексте инструкции, обведенные в рамку. Выполнив инструкции, вы освоите основные этапы настройки программы и обработки документов, сможете увидеть результаты обработки заполненной анкеты. Если при настройке проекта у вас возникнут трудности, вы можете открыть готовый проект Questionnaire.fcproj, файл которого вы найдете в папке <имя диска>:\Documents and Settings\All Users\Application Data\ABBYY\FlexiCapture\8.0\Samples\FlexiCapture-Russian\Questionnaire (для ОС Microsoft Windows Vista <имя диска>:\Users\Public\ABBYY\FlexiCapture8.0\Samples\FormDesigner-Russian\Questionnaire).

Эти и другие примеры проектов доступны из меню Старт > Программы > ABBYY FlexiCapture 8.0 Professional > Справки > Примеры проектов FlexiCapture.

Пожалуйста, пришлите нам проэкспортированные данные в формате .xml. Это поможет нам узнать больше о наших пользователях с целью улучшения продукта.

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
2.	Функции администратора и оператора.....	5
3.	Какие документы можно обрабатывать с помощью ABBYY FlexiCapture 8.0 Professional.....	6
4.	Настройка системы для ввода жестких форм .....	7
4.1.	Создание макета жестких форм.....	8
4.1.1.	Элементы формы .....	9
4.1.2.	Разметка полей данных .....	10
4.2.	Создание проекта .....	12
4.3.	Создание шаблона документа.....	13
4.3.1.	Редактор шаблона документа .....	14
4.3.2.	Геометрическая разметка объектов шаблона.....	15
4.3.2.1.	Группы полей .....	17
4.3.2.2.	Поля без разметки.....	17
4.3.2.3.	Разметка таблиц .....	18
4.3.2.4.	Поля с несколькими областями.....	18
4.3.2.5.	Поля с несколькими экземплярами.....	18
4.3.2.6.	Исключение региона из распознавания.....	19
4.3.2.7.	Удаление полей.....	19
4.3.3.	Статические элементы.....	20
4.3.3.1.	Особенности штрих-кода.....	21
4.3.4.	Свойства полей.....	21
4.3.4.1.	Общие свойства полей.....	21
4.3.4.2.	Типы данных .....	23
4.3.4.2.1.	Типы данных текстового поля ввода .....	23
4.3.4.2.2.	Типы данных для меток .....	25
4.3.4.2.3.	Типы данных для групп меток.....	26
4.3.4.3.	Свойства распознавания полей.....	27
4.3.4.3.1.	Свойства распознавания текстового поля ввода.....	28
4.3.4.3.2.	Свойства распознавания меток и групп меток.....	29
4.3.4.3.3.	Свойства распознавания штрих-кода.....	30
4.3.4.3.4.	Свойства распознавания картинки.....	31
4.3.4.4.	Параметры верификации.....	31
4.3.4.5.	Параметры экспорта картинок.....	32
4.3.4.6.	Проверка с помощью правил.....	33
4.3.5.	Как создать шаблон для многостраничных документов.....	35
4.3.6.	Как создать шаблон для документов с приложениями .....	38
4.3.7.	Настройка экспорта данных.....	38
4.3.7.1.	Экспорт в файл.....	38
4.3.7.2.	Экспорт в базу данных .....	39
4.3.7.3.	Экспорт в библиотеку документов SharePoint.....	41
4.3.7.4.	Экспорт, описанный с помощью скриптовой процедуры .....	43
4.3.8.	Настройка отображения распознанных данных .....	44
4.3.9.	Редактирование и публикация шаблона .....	45
4.3.10.	Тестирование шаблона .....	45
4.4.	Настройка импорта изображений.....	46

5.	Настройка системы для ввода гибких документов.....	48
6.	Особенности ввода неструктурированных документов.....	48
7.	Работа оператора с настроенным проектом.....	49
7.1.	Добавление изображений.....	50
7.2.	Распознавание.....	51
7.3.	Верификация.....	51
7.4.	Экспорт.....	56
8.	Заключение.....	56

# 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 любого предприятия. 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 ввода 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 8.0 Professional.

## 1.2. Автоматизация ввода данных и документов

ABBYY FlexiCapture 8.0 Professional – программное решение для потокового ввода данных с документов различного типа – структурированных (жестких форм), частично структурированных (гибких документов) и неструктурированных документов.

Автоматизированный ввод данных состоит из нескольких 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 каком-либо графическом формате, а также в PDF с возможностью поиска.

Использование системы ABBYY FlexiCapture 8.0 Professional является эффективным решением задачи автоматизации ввода, позволяющее легко управлять процессами, контролировать качество вводимых данных.

Каковы преимущества такого решения?

- **Сокращение затрат.** Систему можно настроить так, что большинство операций будут выполняться автоматически, без участия человека. Оператору нужно будет только складывать пачки документов в лоток сканера и проверять неуверенно распознанные символы. Таким образом, вместо нескольких рабочих мест при ручном вводе, достаточно организовать одно, оборудованное системой ABBYY FlexiCapture 8.0 Professional.
- **Увеличение скорости ввода.** Автоматизация всего процесса заметно увеличивает скорость обработки документов.
- **Высокое качество распознавания.** Благодаря уникальным технологиям ABBYY распознавание символов производится не только быстро, но и качественно. Сопоставимого качества очень трудно добиться при ручном вводе, особенно без снижения скорости работы.
- **Удобство работы и легкость освоения.** ABBYY FlexiCapture 8.0 Professional обладает удобным интерфейсом, как для администратора системы, так и для оператора, обрабатывающего документы. Не требуется длительного обучения сотрудников. Освоить работу с программой легко, в этом вам помогут наши справочные материалы.
- **Масштабируемость.** Для увеличения производительности вам нужно установить ABBYY FlexiCapture 8.0 Professional на еще одно рабочее место. В отличие от процесса ручного ввода, масштабировать автоматизированную систему достаточно легко.

## 2. Функции администратора и оператора

Для организации потокового ввода данных с помощью ABBYY FlexiCapture 8.0 Professional нужно настроить систему для обработки определенного вида документов. Все необходимые настройки выполняет администратор, а ввод документов осуществляется операторами. Количество рабочих мест определяется в зависимости от числа и объема обрабатываемых документов.

**Администратор** выполняет настройку системы и контролирует ее работу. Обязанностями администратора являются:

- разработка бланков новых форм (это может входить в обязанности дизайнера);
- настройка импорта изображений, параметров сканирования;
- подготовка шаблонов документов, которая включает:
  - задание параметров распознавания;
  - определение правил проверки и сборки документов;
  - настройку экспорта данных;
- мониторинг процесса ввода данных в настроенной системе, консультирование операторов.

**Операторы** выполняют ввод документов. Они работают с уже настроенной администратором системой, готовой к обработке заполненных документов определенного вида. В обязанности операторов входит:

- сканирование или импорт документов;
- верификация данных;

- экспорт данных.

Распознавание документов производится автоматически.

В программе предусмотрены два режима работы: режим администратора и режим оператора. Администратор имеет доступ ко всем функциям программы, режим администратора может быть защищен паролем. Оператор не имеет доступа к созданию и редактированию шаблонов, созданию профилей импорта. Оператору можно разрешить работу только с одним проектом или только с одним пакетом.

При установке программы можно установить только станцию оператора или выбрать полную установку, в этом случае можно будет работать как в режиме администратора, так и в режиме оператора. Вы можете запустить программу в одном из этих режимов, выбрав соответствующий пункт меню **Пуск > Программы > ABBYY FlexiCapture 8.0 Professional > ABBYY FlexiCapture 8.0 Станция администратора** или **Пуск > Программы > ABBYY FlexiCapture 8.0 Professional > ABBYY FlexiCapture 8.0 Станция оператора**.

### 3. Какие документы можно обрабатывать с помощью ABBYY FlexiCapture 8.0 Professional

ABBYY FlexiCapture 8.0 Professional – программное решение для потокового ввода данных с документов различного типа.

Различные документы могут обрабатываться в одном потоке, а можно настроить обработку документов смешанного типа. Тип документов влияет только на настройку системы, а именно, на способ создания шаблона документов. Работа оператора не зависит от типа обрабатываемых документов.

Рассмотрим подробнее разные виды документов, которые вы можете обрабатывать с помощью ABBYY FlexiCapture 8.0 Professional.

- **Структурированные документы.** К структурированным относятся документы с набором выделенных информационных полей, оформление, количество и местоположение которых не меняются от одного экземпляра документа к другому. Такие документы называются жесткими формами. Например, анкеты, опросники, заявления чаще всего являются жесткими формами. Чаще всего такие формы распространяются в виде бланков и заполняются от руки. Чтобы идентифицировать жесткую форму в потоке документов и извлечь из нее данные, необходимо создать единую жесткую разметку, которая укажет программе места расположения полей, из которых нужно извлечь данные. Наиболее эффективно производится обработка жестких форм, созданных с учетом требований автоматического ввода. Такие формы называют машиночитаемыми формами. В состав программы входит удобный инструмент разработки машиночитаемых форм ABBYY FormDesigner 8.0 (Поставляется вместе с ABBYY FlexiCapture 8.0 Professional). О создании форм с помощью ABBYY FormDesigner 8.0 читайте в руководстве и справке к этой программе. Основные этапы создания шаблона описаны именно для структурированных документов.

- **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 8.0 Professional использует гибкую разметку (FlexiLayout). Для создания гибкой разметки используется специальный модуль 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 8.0 Professional. 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.

Далее описаны основные принципы настройки ABBYY FlexiCapture 8.0 Professional в зависимости от вида обрабатываемых документов. Вы узнаете об этапах процесса автоматизированного ввода данных, получите представление о том, как можно улучшить качество распознавания и как организовать экспорт данных.

Наиболее подробно описан ввод структурированных документов – жестких форм, на примере которых мы расскажем обо всех этапах процесса, а для других видов документов опишем их особенности и отличия при создании шаблона.

## 4. Настройка системы для ввода жестких форм

ABBYY FlexiCapture 8.0 Professional позволяет эффективно вводить жесткие формы, их обработка производится просто и качественно. Процесс работы с такими формами состоит из следующих этапов:

- разработка бланков жестких форм или приведение уже используемых форм к машиночитаемому виду;
- создание шаблона: геометрическая разметка объектов, задание их свойств, создание правил проверки правильности введенных значений и правил сборки страниц в документы, настройку экспорта данных;
- настройка способов добавления изображений, создание профилей импорта изображений;

- и, наконец, ввод документов, к которому можно приступить, когда все необходимые настройки выполнены.

## 4.1. Создание макета жестких форм

Разработка макета бумажной формы – очень важный этап процесса, т.к. от вида формы зависит удобство ее заполнения и возможность безошибочного автоматизированного ввода. Человеку, заполняющему форму, должно быть понятно, куда и какие данные вносить, в этом случае снижается вероятность возникновения ошибок. При разработке макета нужно стремиться к тому, чтобы процесс заполнения создаваемой формы был интуитивно понятен, а введенная информация легко распознавалась.

Форма, разработанная с учетом требований автоматизированного ввода, называется **машиночитаемой формой**. Такая форма обладает следующими свойствами:

- Форма имеет **реперы** – служебные элементы, с помощью которых программа определяет ориентацию формы, производит наложение на нее шаблона и выравнивает искажения изображения, полученные при сканировании. Реперами могут быть черные квадраты, кресты или уголки.
- Все поля и графические элементы расположены одинаково на каждом экземпляре формы. Формы одного типа должны быть отпечатаны с одного макета и совпадать на просвет.
- Пояснительная информация расположена так, чтобы она не мешала извлечению содержимого из полей для ввода данных.

Вы можете подробнее ознакомиться с требованиями к машиночитаемым формам в справке ABBYY FlexiCapture 8.0 Professional и в руководстве по созданию машиночитаемых форм.

Кроме того, для создания машиночитаемых форм компания ABBYY разработала удобный инструмент – ABBYY FormDesigner 8.0, входящий в комплект поставки ABBYY FlexiCapture 8.0 Professional. (О создании машиночитаемых форм с помощью ABBYY FormDesigner читайте в справке и руководстве FormDesigner.) С помощью программы ABBYY FormDesigner вы легко освоите создание форм, удовлетворяющих требованиям машиночитаемости, кроме того, шаблон для обработки форм, созданных с помощью ABBYY FormDesigner, можно импортировать в программу ABBYY FlexiCapture 8.0 Professional, получив готовую разметку полей и статических элементов. Вам останется только задать свойства полей и настроить экспорт.

Для того чтобы привести к машиночитаемому виду уже используемые вами формы, достаточно немного изменить дизайн и добавить некоторые элементы (в частности, реперы). Если же вы по каким-то причинам не можете изменить бланки форм, то вы сможете настроить программу и для обработки форм, не имеющих реперов. В качестве реперов можно использовать некоторые другие элементы формы, например, вертикальные и горизонтальные линии, пояснительный текст или штрих-код, однако наложение шаблона на форму со стандартными реперами производится наиболее эффективно.



### 4.1.1. Элементы формы

Рассмотрим основные *элементы формы*. (Рисунок 1)

Заголовок формы

Реперы

Поля для ввода текста

Метки

Пояснительный текст

Поле для ручного ввода

Штрих-код

АНКЕТА ПОЛЬЗОВАТЕЛЯ

Дата заполнения: / /  -/ /

Фамилия и Имя:

Должность:

Адрес электронной почты:

Название компании:

Страна:

Индустрия:

Государственные службы <input type="checkbox"/>	Финансы/Банки <input type="checkbox"/>	Здравоохранение <input type="checkbox"/>	Страхование <input type="checkbox"/>
Информационные технологии <input type="checkbox"/>	Образование <input type="checkbox"/>	Сервисные бюро <input type="checkbox"/>	Производство <input type="checkbox"/>

Другое:

Тип обрабатываемых документов:  
(можно выбрать несколько)

Чеки <input type="checkbox"/>	Тесты <input type="checkbox"/>	Заявления <input type="checkbox"/>	Анкеты <input type="checkbox"/>
Платежные документы <input type="checkbox"/>	Паспорта <input type="checkbox"/>	Счета-фактуры <input type="checkbox"/>	Заказы <input type="checkbox"/>

Другое:

Объем обрабатываемых документов (страниц в месяц):

Как обрабатываются документы в Вашей компании в настоящее время:

Не обрабатываются <input type="checkbox"/>	Данные вводятся вручную <input type="checkbox"/>	Сканируются <input type="checkbox"/>	FormReader 6.5 <input type="checkbox"/>
--	--	--------------------------------------	---

Другое:

Производитель сканера, который Вы планируете использовать для ввода документов:  
(можно выбрать несколько)

Konica Minolta <input type="checkbox"/>	Fujitsu <input type="checkbox"/>	Kodak <input type="checkbox"/>	Xerox <input type="checkbox"/>	HP <input type="checkbox"/>	Kyocera Mita <input type="checkbox"/>
---	----------------------------------	--------------------------------	--------------------------------	-----------------------------	---------------------------------------

Другое:

В какое электронное хранилище Вы планируете экспортировать обработанные данные и документы?  
(можно выбрать несколько)

MS Access <input type="checkbox"/>	MS SQL Server <input type="checkbox"/>	Documentum <input type="checkbox"/>	IBM Content Manager <input type="checkbox"/>	MS SharePoint <input type="checkbox"/>
------------------------------------	--	-------------------------------------	--	--

Другое:

Ваши комментарии и пожелания:

Я хочу поделиться опытом использования ABBYY FlexiCapture. Свяжитесь со мной по e-mail

Спасибо за заполнение анкеты! Пожалуйста, обработайте анкету с помощью ABBYY FlexiCapture и отправьте данные по адресу [FlexiCapture\\_feedback@abbyy.com!](mailto:FlexiCapture_feedback@abbyy.com)

Заполняя эту форму, вы тем самым добровольно передаете нам свои персональные данные и таким образом выражаете свое согласие на сбор, обработку и использование этих персональных данных на условиях обеспечения конфиденциальности и в соответствии с ФЗ РФ «О персональных данных». Представленные Вами персональные данные будут использованы только в рамках группы компаний ABBYY и не будут предоставлены третьим лицам, за исключением случаев, предусмотренных действующим законодательством.

© 2007 ABBYY.  
SharePoint, SQL Server являются зарегистрированными товарными знаками Корпорации Microsoft в Соединенных Штатах и/или других странах. Xerox – товарный знак Корпорации Xerox в США и/или других странах. Kodak – товарный знак Eastman Kodak Компании. IBM – зарегистрированный товарный знак IBM в США и/или других странах. Все упомянутые товарные знаки являются собственностью их владельцев.

Рисунок 1. Пример машиночитаемой формы с основными элементами

© 2009 ABBYY. Все права защищены.

- **Поля данных.** Любая форма, предназначенная для сбора информации, содержит поля для внесения данных, которые обычно сопровождаются пояснительными надписями. Существуют поля для ввода информации следующих видов:

**Текстовые поля** предназначены для ввода текста. Они представляют собой группу знакомест для заполнения символами. Дизайн текстового поля побуждает заполняющего писать символы отдельно.

**Метки** также предназначены для внесения информации, но в них нужно не писать текст, а отмечать соответствующие пункты. Метка обычно выглядит как замкнутый контур (квадрат, круг, многоугольник), информация в который вносится путем постановки условного знака (например, галочки или крестика) внутри контура. В некоторых случаях метка может не иметь контура, а условный знак проставляется на белом поле в отведенном для него месте. Если вы хотите допускать исправления меток, вы сможете отметить соответствующую опцию при создании шаблона, тогда полностью закрашенная метка будет считаться неотмеченной.

**Группы меток.** Несколько меток, расположенных рядом и объединенных по смыслу, образуют группу. Как правило, метки внутри одной группы соответствуют взаимоисключающим вариантам ответа, то есть в группе должна быть выбрана только одна метка.

Поля для ввода данных могут также быть представлены в виде таблиц.

- **Реперы.** Для определения ориентации страницы и точного наложения шаблона нужны блоки-реперы, с помощью которых программа также отслеживает и корректирует искажения в изображении формы, полученные при сканировании. Реперы могут выглядеть как черные квадраты (предпочтительно), уголки, кресты. Рекомендуется размещать на форме 5 реперов: четыре – по углам формы и один – на одной из сторон, для уверенного определения ориентации страницы. В этом случае наложение шаблона будет точным и быстрым, кроме того, вы сможете вводить в едином потоке формы, напечатанные на разных принтерах или переданные по факсу.
- **Идентификаторы.** Они служат для определения принадлежности страницы к той или иной форме и для выбора шаблона при наличии в пакете нескольких шаблонов со сходным набором реперов. В случае обработки нескольких форм в одном потоке необходимо на каждой странице формы предусмотреть уникальный элемент, указывающий на принадлежность страницы к той или иной форме. В качестве такого элемента можно использовать штрих-код, реперы, разделители или статический текст (название формы или пояснительный текст).
- **Графические изображения.** Вам может понадобиться сохранять некоторые объекты в виде изображений, например, фотографию, подпись, печать, штамп. При помощи ABBYY FlexiCapture 8.0 Professional можно сохранять изображения объектов и помещать их в файл или базу данных.
- **Декоративные элементы.** Форма может содержать элементы оформления, например, логотип компании.

#### 4.1.2. Разметка полей данных

По способу разметки и выделения знакомест машиночитаемые формы делятся на три типа: цветные фоновые, растровые и черно-белые линейчатые формы.

- **Цветные фоновые формы.** Знакоместа на такой форме выполнены в виде белых прямоугольников на цветном или сером фоне. Каждое знакоместо предназначено для написания одного символа. Цвет и насыщенность фона нужно подбирать таким образом, чтобы его можно было удалить при сканировании. В идеальном случае после сканирования с формы должны исчезнуть все элементы, кроме реперов и заполненных полей ввода. Для подобной обработки используют либо монохромные сканеры с цветной (красной или зеленой) лампой, либо цветные сканеры с настройками, позволяющими удалить цвет фона.

**Дата заполнения:**  /  /

**Фамилия и Имя:**

**Должность:**

**Адрес электронной почты:**

- **Растровая форма.** На такой форме для ограничения знакомест используются точки, расположенные на одинаковом расстоянии друг от друга. После сканирования точки удаляются с изображений без потери информации в заполненных полях. Для этого включается настройка «Удаление мусора». Знакоместа могут быть изображены в виде прямоугольников, ограниченных линиями из точек, либо в виде белых прямоугольников на состоящем из точек фоне.

**Дата заполнения:**  /  /

**Фамилия и Имя:**

**Должность:**

**Адрес электронной почты:**

- **Черно-белая линейчатая форма.** Границы информационных полей и знакомест на такой форме задаются обычными линиями, которые не исчезают при сканировании. Отделение содержания полей от разметки происходит в процессе распознавания. Программа находит вертикальные и горизонтальные линии в разметке и отделяет их от введенных данных на основе информации о том, из скольких ячеек состоит поле, и как ячейки отделены друг от друга.

**Дата заполнения:**  /

**Фамилия и Имя:**

**Должность:**

**Адрес электронной почты:**

**Анкета, которую вы получили с программой ABYU FlexiCapture 8.0 Professional – пример жесткой машиночитаемой формы. Легко определить, что это фоновая форма. Посмотрите, как расположены на ней основные элементы.**

## 4.2. Создание проекта

**Проект** включает в себя все необходимые настройки для потокового ввода документов: шаблоны документов, профили импорта изображений, настройки программы и обрабатываемые документы.

Документы объединяются в **пакеты**. Количество пакетов зависит от принципа обработки: вы можете обрабатывать все документы в рамках одного пакета, а можете объединять в пакеты документы, введенные одной и той же датой или из отсканированного за раз набора документов.

Для обработки документов используются **рабочие пакеты**. В режиме оператора доступны только они. **Тестовые пакеты** используются администратором в процессе отладки шаблона. Отличие этих пакетов в том, что для обработки документов из тестовых пакетов используется локальная версия шаблона, а для обработки документов из рабочих пакетов – опубликованная версия.

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

В проект можно включать несколько шаблонов, 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.

Для начала работы администратору нужно создать проект и хотя бы один шаблон документов.

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 (Рисунок 2). 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**.

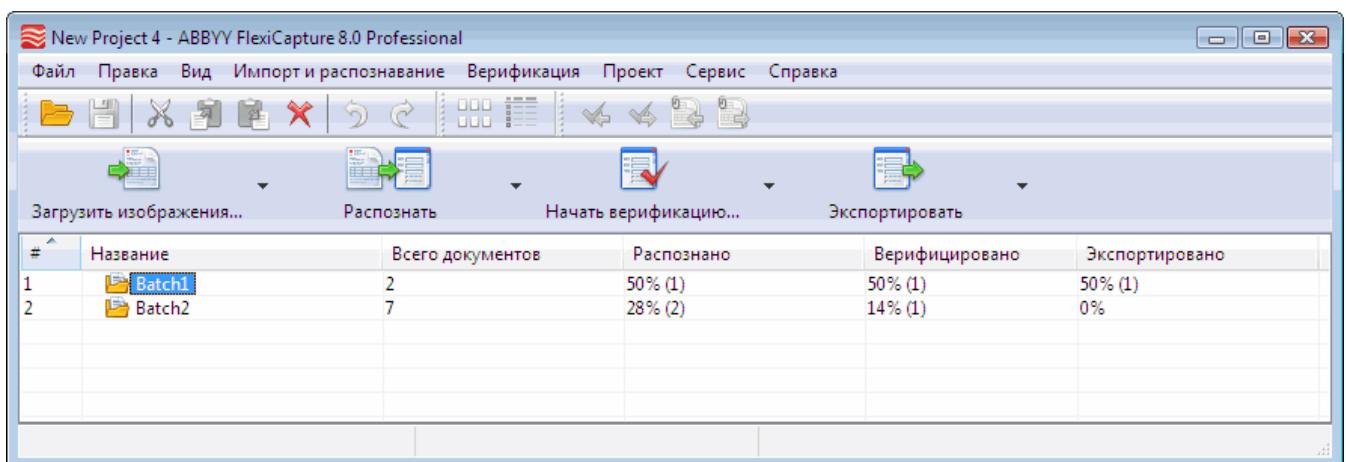


Рисунок 2. Главное окно программы ABBYY FlexiCapture 8.0 Professional

Откройте программу ABBYY FlexiCapture 8.0 Professional. Для создания нового проекта нажмите кнопку **Новый...** в диалоге *Открытие проекта*, либо выберите пункт меню *Файл>Новый проект...* в основном окне программы. Определите директорию для хранения проекта и задайте его имя. При нажатии кнопки *Создать* откроется новый проект.

### 4.3. Создание шаблона документа

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

- отметить на изображении **статические элементы**: реперы, разделители, статический текст, штрих-коды, и определить, какие из них будут служить для наложения шаблона и для идентификации. Нахождение и выделение реперов производится автоматически.
- указать местоположение всех **полей**. Поля должны соответствовать областям изображения, из которых нужно извлечь данные.
- 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 или базу данных, или согласно скриптовой процедуре.

После того как шаблон будет создан, его нужно опубликовать, чтобы он стал доступен для распознавания документов.


Для создания нового шаблона выберите пункт меню **Проект > Шаблоны документов...** и в появившемся диалоге нажмите кнопку **Новый...** Откроется Мастер создания нового шаблона документов. В диалоге **Создание нового шаблона документа** вы можете задать основные свойства шаблона: 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 шаблон. (Если ваш документ состоит из нескольких страниц, то загрузите первую страницу, а для добавления остальных руководствуйтесь рекомендациями раздела Как создать шаблон для многостраничных документов). Вы можете отсканировать изображение незаполненной страницы или загрузить его из файла. Если вы собираетесь обрабатывать полуструктурированные документы, то вам нужно использовать гибкое описание для создания шаблона. 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).

На следующем шаге 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 их обычно отмечают вручную.

Если на форме есть реперы, то они также будут найдены и отмечены на изображении автоматически.

1. В основном окне программы выберите пункт меню **Проект > Шаблоны документов...** Нажмите кнопку **Новый...** в диалоге **Шаблоны документов**.
2. В открывшемся диалоге введите параметры шаблона: его название и комментарий. В поле **Язык(регион)** следует выбрать тот язык, на котором вы будете заполнять анкету. В поле **Стиль написания** выберите вашу страну, т.к. часто стиль написания одних и тех же символов, например цифр, отличается в разных странах. Выберите **ICR (рукопечатный)** тип текста. Нажмите кнопку **Далее (Next)**.
3. На следующем шаге выберите опцию **Сканер** и отсканируйте незаполненную анкету без применения фильтров, чтобы цвет фона не исчез при сканировании. Так будет легче разметить поля шаблона на изображении формы. (Позже, когда вы будете сканировать изображения страниц для распознавания данных, вы выберете режим сканирования, при котором фон пропадает). На основе этого изображения будет создан шаблон страницы. Вы также можете отметить опцию **Загрузить из файла** и указать путь к изображению. (Изображение анкеты вы найдете в папке <имя диска>:\Documents and Settings\All Users\ Application Data\ABBYY\FlexiCapture\8.0\Samples\FormDesigner\Russian\Questionnaire, для ОС Microsoft Windows Vista <имя диска>:\Users\Public\ABBYY\FlexiCapture8.0\Samples\FormDesigner\Russian\Questionnaire). Нажмите кнопку **Далее (Next)**.
4. На следующем шаге можно отметить опции **Текст** и **Метки** для автоматического выделения полей для ввода текста и меток. Нажмите **Готово (Finish)**.
5. После завершения работы мастера откроется окно редактора шаблона с изображением анкеты, на котором будут выделены поля для ввода текста и метки. Названия автоматически выделенных полей шаблона отобразятся в окне **Структура документа**. Перейдя в режим статических элементов (кнопка ) вы увидите разметку статических элементов – реперы также выделены автоматически.

### 4.3.1. Редактор шаблона документа

Основные действия по созданию и редактированию шаблона производятся в окне **редактора шаблона документа** (Рисунок 3), которое открывается после завершения работы мастера создания нового шаблона. Для того чтобы открыть редактор шаблона документа из главного окна, нужно выбрать пункт меню **Проект > Шаблоны документов...**, выделить название шаблона и нажать кнопку **Правка....**

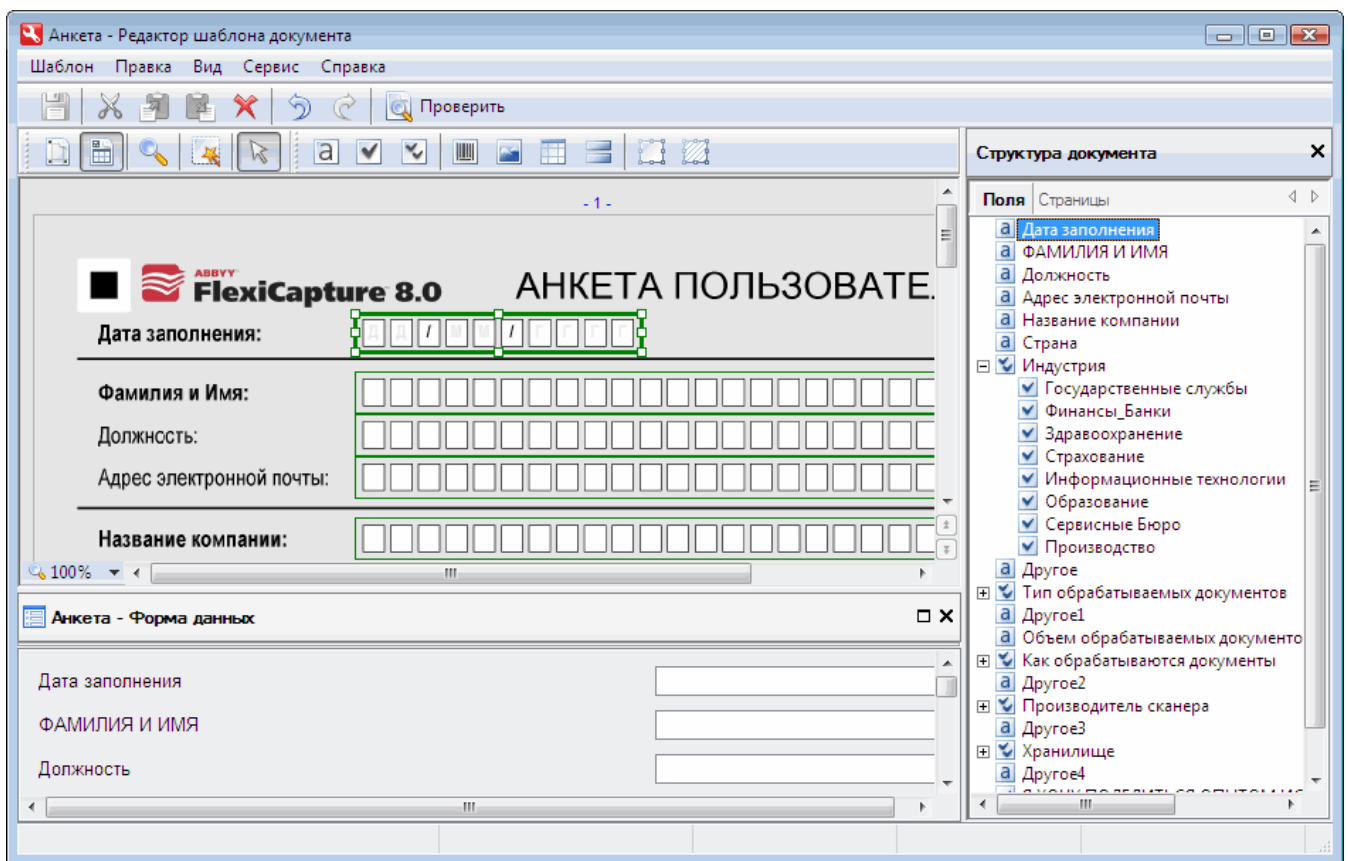





Рисунок 3. Редактор шаблона документа

### 4.3.2. Геометрическая разметка объектов шаблона

После прохождения всех шагов мастера создания шаблона в окне редактора шаблона 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 будет определено автоматически.

Редактор шаблона документа снабжен набором удобных и наглядных инструментов для разметки полей и статических элементов. В редакторе существуют два режима:

- 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



- метка



- группа меток



- штрих-код



- графический элемент (картинка)



- таблица



- группа полей

### Статические элементы:



- репер



- разделитель



- статический текст



- штрих-код

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 шаблона, 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. 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 пояснительный 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 .

**1. Поля для ввода текста уже выделены на форме автоматически. Если вы не отметили опцию *Текст* на последнем шаге мастера создания шаблона или по другим причинам хотите выделить поля для ввода**

**текста вручную, выберите инструмент  и выделите все поля, в которые будет внесен текст. Поле должно включать в себя все знакоместа с небольшим запасом со всех сторон. Обведите таким образом поля для ввода даты, имени, должности, страны, адреса электронной почты, имени компании, количества страниц, обрабатываемых в день, 4 поля для ввода дополнительной информации – Другое.**


**2. Отметьте также поле для ввода пожеланий и комментариев Ваши комментарии и пожелания. Это поле не было найдено автоматически, так как не имеет разметки.**



3. Если вы не выделили метки автоматически, выберите инструмент  и обведите все квадраты для меток, не входящих в группы, также оставляя небольшой запас. Метки, входящие в группу, не обязательно отмечать по отдельности. Достаточно выбрать  и обвести прямоугольником все метки, которые должны входить в группу. Каждая метка будет выделена автоматически, ей присваивается имя, и все эти метки объединяются в группу.
4. Перейдите в режим статических элементов. Реперы уже отмечены на вашей форме. Выберите инструмент  и выделите штрих-код на изображении.

#### 4.3.2.1. 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.


#### 4.3.2.2. Поля без разметки

Поля могут не иметь соответствующей области на изображении. Названия полей без разметки в списке помечены красной звездочкой. Такие поля могут использоваться для хранения промежуточных 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 шаблона выберите пункт меню **Шаблон > Создать поле** и создайте поле нужного типа. Название поля 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 название поля 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.)


#### 4.3.2.3. Разметка таблиц

ABBYY FlexiCapture 8.0 Professional предоставляет возможность работы с таблицами. Создание полей типа **Таблица** существенно облегчает настройку, извлечение данных и экспорт данных из таблиц. Поле **Таблица** представляет собой набор из столбцов одного типа и повторяющихся строк.

Для разметки таблиц в жестком шаблоне 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. После того, как на изображении будет создана таблица, на панели инструментов появятся инструменты для разметки таблиц. 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.

#### 4.3.2.4. 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.

#### 4.3.2.5. 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 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.

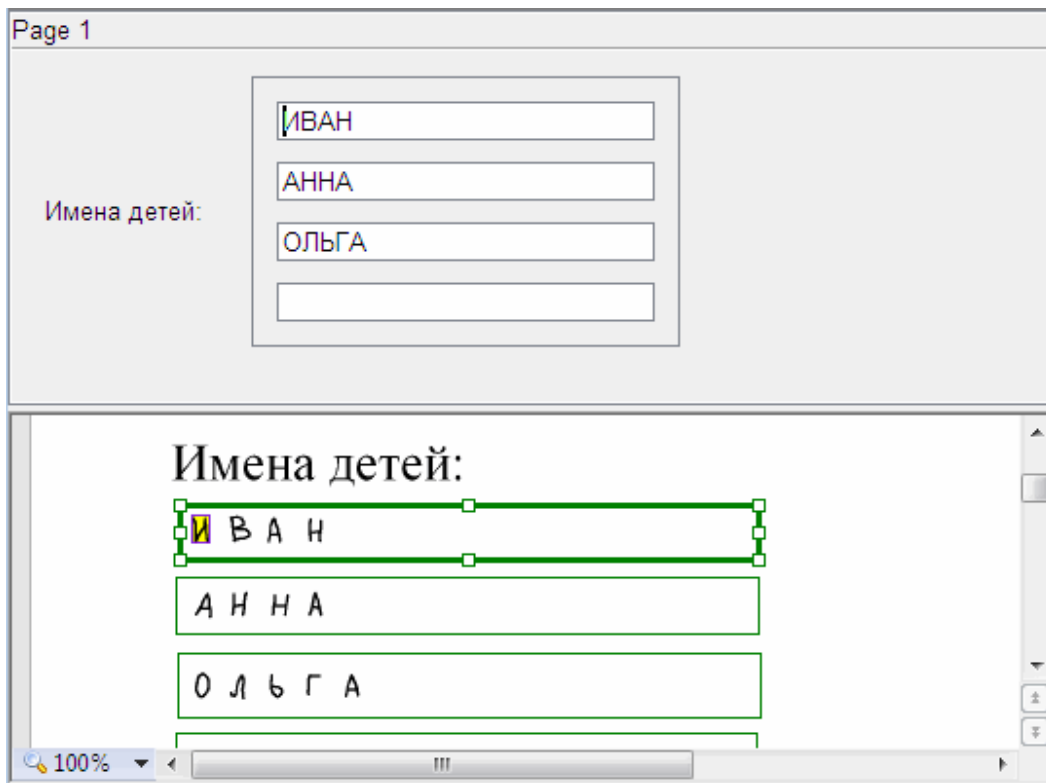


Рисунок 4. Пример использования полей с несколькими экземплярами

#### 4.3.2.6. Исключение региона из распознавания

Иногда бывает необходимо исключить из распознавания некоторую область, например, область с пояснительным текстом, который мешает извлечению данных из поля (Рисунок 24). Чтобы


исключить регион из распознавания, выберите инструмент  и выделите с помощью мыши область, которую нужно исключить.



Рисунок 5. Пример использования нераспознаваемого региона


#### 4.3.2.7. 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.

### 4.3.3. Статические элементы

Статические элементы – это объекты, из которых не извлекается информация. Они служат для наложения шаблона и идентификации.

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

Для наложения шаблона 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 шаблона 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.

**Репер**: отметьте опцию **Использовать для наложения шаблона**, если репер будет использоваться для наложения шаблона (рекомендуется), и опцию **Использовать для идентификации шаблона** – для идентификации. Задайте тип репера (квадрат, крест, уголок, прямоугольник).

**Статический текст**: отметьте опцию **Использовать для наложения шаблона**, если текст будет использоваться для наложения шаблона, и опцию **Использовать для идентификации шаблона** – для идентификации. 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**: отметьте опцию **Использовать для наложения шаблона**, если разделитель будет использоваться для наложения шаблона, и опцию **Использовать для идентификации шаблона** – для идентификации.

**Штрих-код**: отметьте опцию **Использовать для наложения шаблона**, если штрих-код будет использоваться для наложения шаблона, и опцию **Использовать для идентификации шаблона** – для идентификации. 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.

На изображении вашей формы отмечено 5 реперов. Убедитесь, что для всех реперов в диалоге свойств выбран тип **Черный квадрат** и отмечены опции **Использовать для наложения шаблона**, **Использовать для идентификации шаблона**.

Для штрих-кода также отметьте опции **Использовать для наложения шаблона** и **Использовать для идентификации шаблона**. Если значение штрих-кода не появилось в поле **Значение** автоматически, то нажмите кнопку **Подставить**. Значение штрих-кода будет подставлено.

Штрих-код на данной форме соответствует типу **EAN 13**, это должно быть отмечено в поле **Тип кода** на закладке **Распознавание**. **Ориентация** – **Слева направо**.

**5 реперов и штрих-код – достаточный набор элементов для однозначного наложения шаблона и его идентификации, если мы не планируем обрабатывать в одном потоке другие документы с аналогичным расположением реперов.**

#### 4.3.3.1. 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.

#### 4.3.4. 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, т.к. проверка правильности значений будет производиться автоматически.

Когда вы создали поле определенного типа, ему присваиваются свойства по умолчанию. Их можно изменить в диалоге, который открывается посредством выбора пункта **Свойства** контекстного меню любого объекта. Поле каждого типа имеет свой набор свойств.

##### 4.3.4.1. Общие свойства полей

Закладка **Общие** (Рисунок 6) 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. 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.
- **Не может быть пустым** – отметьте эту опцию, если поле должно быть обязательно заполнено.
- **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.

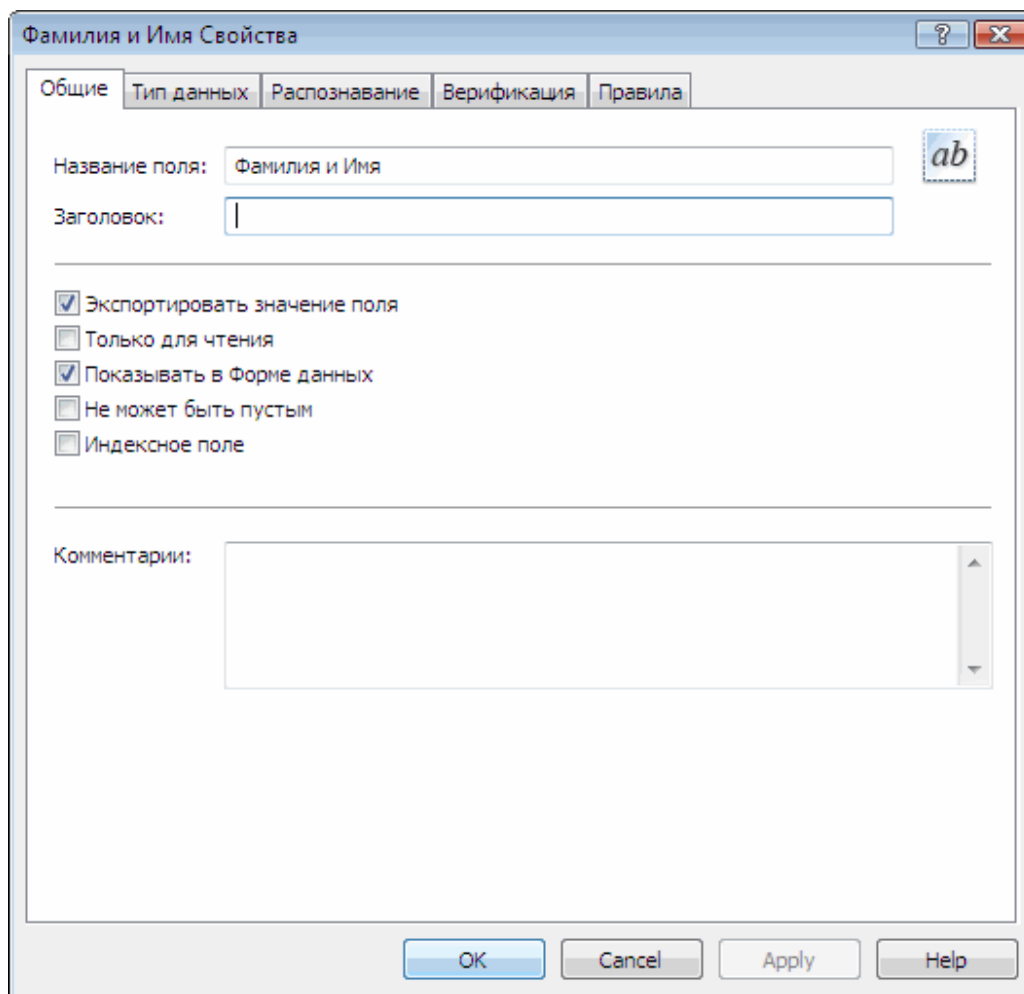


Рисунок 6. Диалог задания свойств поля, закладка для задания основных опций (для текстового поля ввода)

При создании полей им автоматически присваиваются названия, совпадающие с пояснительным текстом. Проверьте, правильно ли названы созданные вами поля. При необходимости измените их названия на закладке **Общие** диалога задания основных свойств поля.

Введите, если нужно, комментарии.

Опции **Экспортировать значение поля** и **Показывать в Форме данных** проставлены по умолчанию для всех полей. Убедитесь, что опция **Экспортировать значение поля** отмечена для всех текстовых полей, меток, не входящих в группы, и для групп меток. (Свойства меток, входящих в группу, являются общими и задаются для всей группы).

#### 4.3.4.2. 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.

##### 4.3.4.2.1. 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.

Тип данных задается на закладке **Тип данных** диалога свойств поля (Рисунок 7).

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.

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 значение поля будет проверяться на вхождение в эти словари;
- 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 предопределенные 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.

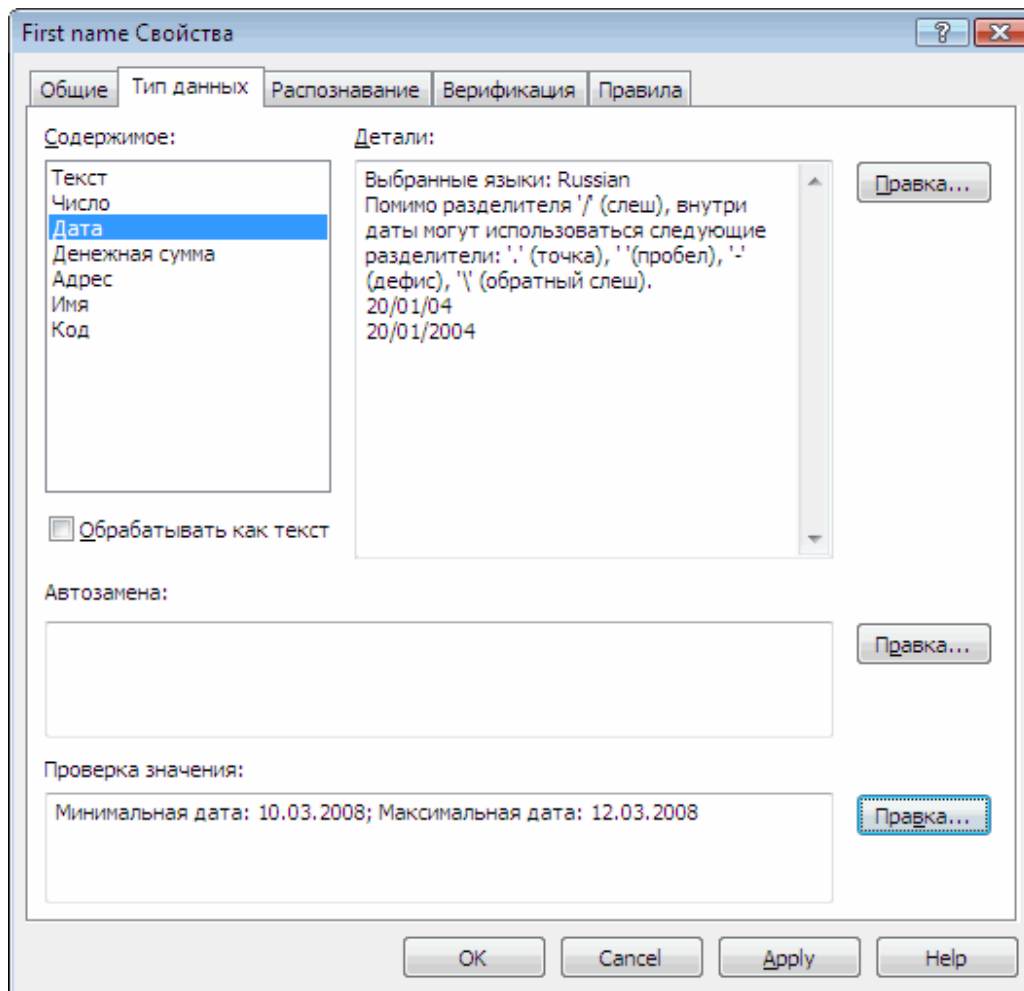


Рисунок 7. Диалог задания свойств поля, закладка для задания типов данных (для текстового поля ввода)

Для любых типов данных можно произвести автоматическую обработку введенного 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 8).



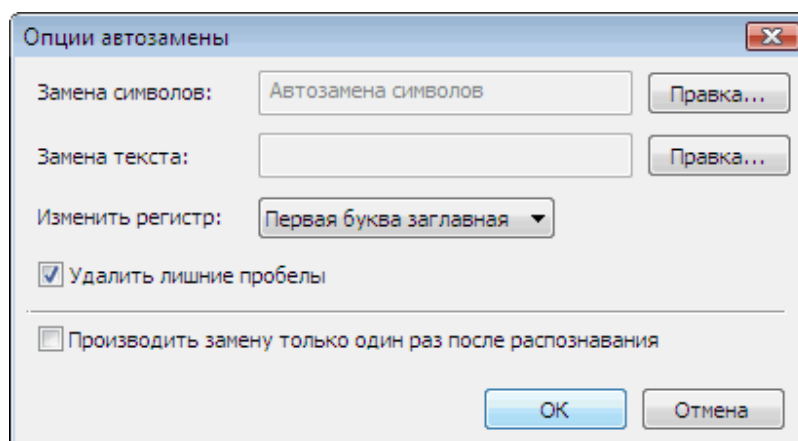


Рисунок 8. Диалог задания параметров автозамены

Также можно настроить проверку распознанного значения поля: проверять, попадает ли значение в заданную область. Чтобы задать ограничение области значений для поля, нажмите кнопку **Правка...** справа от поля **Проверка значения**. (Рисунок 9).

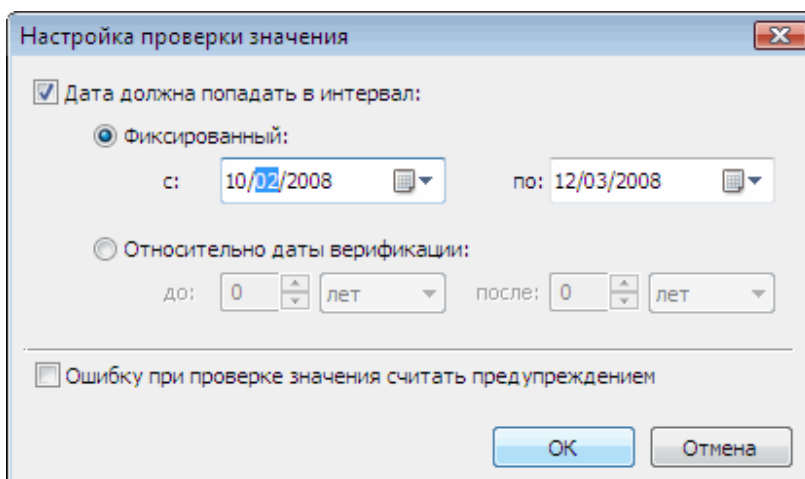


Рисунок 9. Диалог настройки проверки значения

Задайте тип данных для каждого текстового поля. Для поля *Фамилия* и *Имя* следует выбрать тип *Имя* и правильно задать язык, а, например, для поля *Объем обрабатываемых документов*, в которое должно вноситься целое число страниц, выбираем тип *Число* (формат – целое число).

Для поля *Дата* заполнения выберите содержимое *Дата*. Затем выберите из списка или опишите формат даты. Для того чтобы осуществлялась проверка попадания даты в заданный интервал времени, нажмите кнопку **Правка...** в области **Проверка значения** и в открывшемся диалоге отметьте опцию *Дата должна попадать в интервал*. Вы можете задать фиксированный или относительный интервал. Например, вы можете задать такой интервал времени: 90 дней до даты верификации и 0 дней после. В этом случае ошибка не будет появляться для анкет с датой заполнения не ранее, чем 90 дней до настоящего момента и не позднее, чем текущая дата.

#### 4.3.4.2.2. Типы данных для меток

Для меток на закладке **Тип данных** определяются значения, которые будут присваиваться полям в случае отмеченных и неотмеченных меток (Рисунок 10).

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.

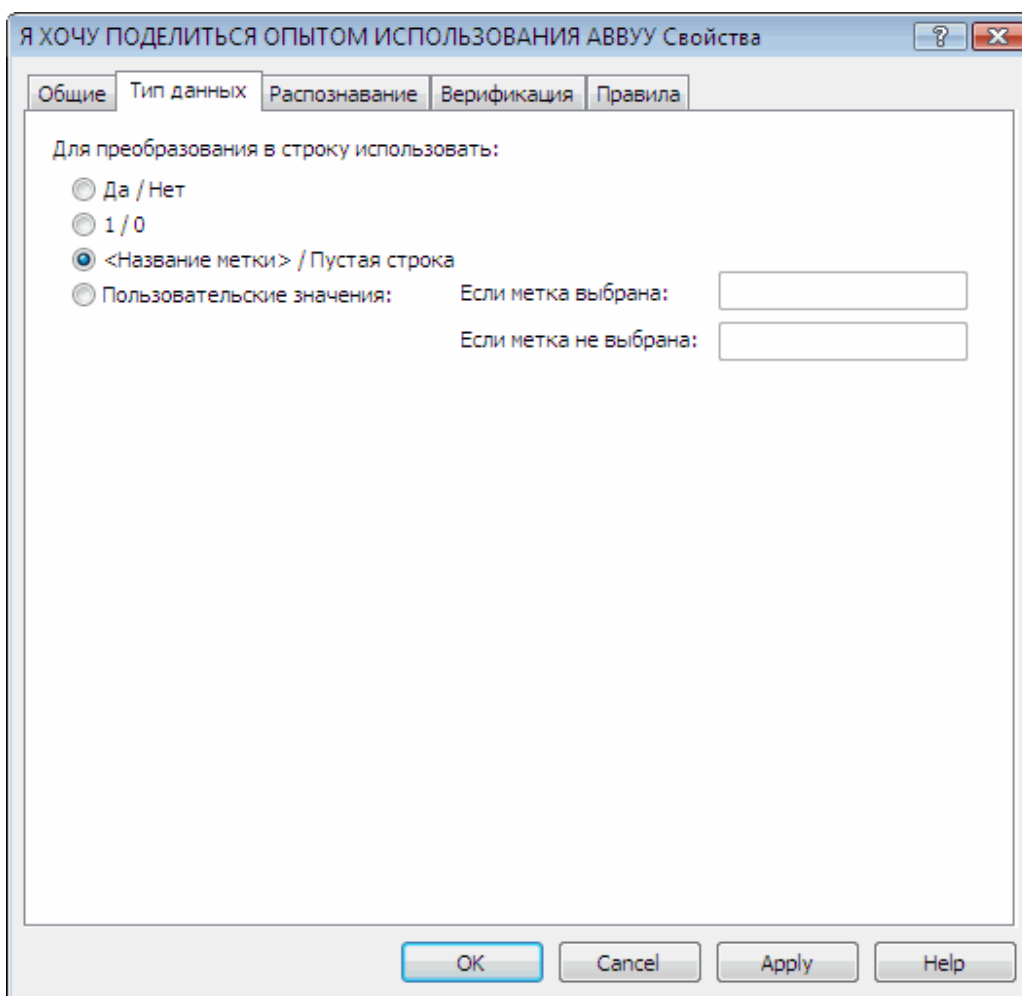


Рисунок 10. Диалог задания свойств поля, закладка для задания типов данных (для метки, не входящей в группу)

#### 4.3.4.2.3. Типы данных для групп меток

В диалоге свойств групп меток на закладке **Тип данных** отображается список названий меток, входящих в данную группу (Рисунок 11).

Отметьте опцию **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.

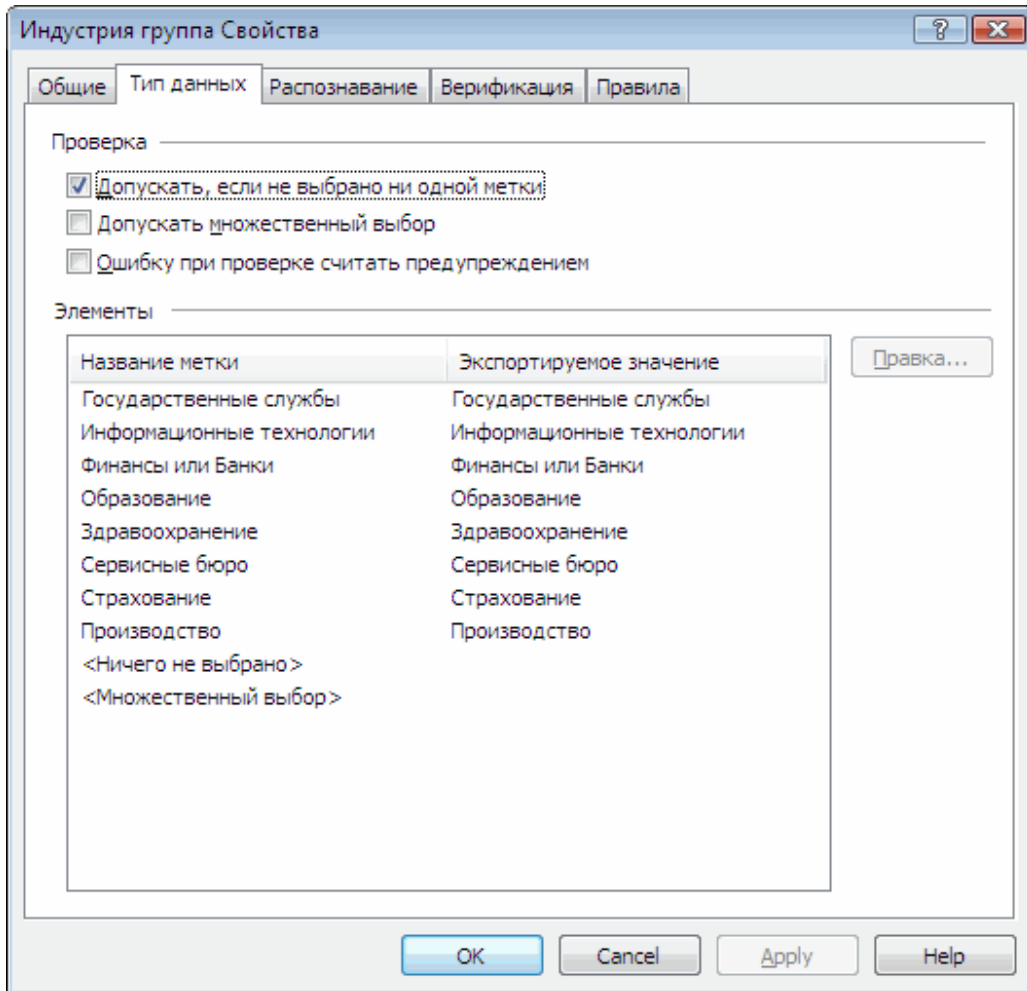


Рисунок 11. Диалог задания свойств поля, закладка для задания типов данных (для групп меток)

**Задайте свойства для меток и групп меток. Для меток выберите способ преобразования значений <Название метки>/Пустая строка.**

**Для группы меток Тип обрабатываемых документов отметьте опции *Допускать, если не выбрано ни одной метки*, и *Допускать множественный выбор*. Задайте экспортируемое значение для случая, если не выбрано ни одной метки, например «ничего не выбрано». Для случая, когда выбрано несколько меток, задайте значение «выбрано несколько меток».**

**Задайте аналогично свойства других групп меток.**

#### 4.3.4.3. Свойства распознавания полей

В программе ABBYY FlexiCapture 8.0 Professional 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.

#### 4.3.4.3.1. *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 машине 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 рекомендуемый тип разметки является моноширинным (*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).

Вы можете выбрать **Регистр** букв, тогда все распознанные буквы будут преобразованы в строчные или заглавные. 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.

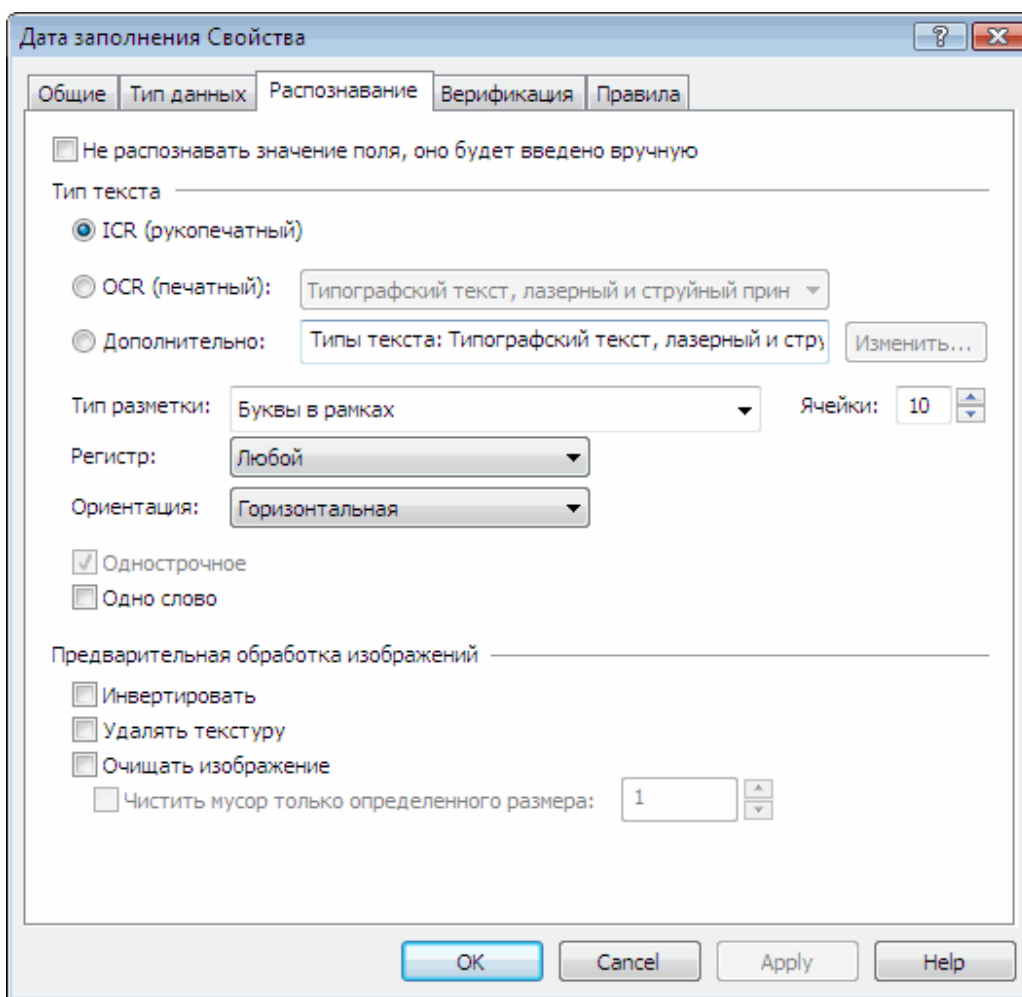


Рисунок 12. Диалог задания свойств поля, закладка для задания опций распознавания (для текстового поля ввода)

1. **Задайте свойства распознавания для всех текстовых полей на вашей форме. Тип текста для всех полей должен быть ICR (рукопечатный), тип разметки – Буквы в изолированных рамках, количество ячеек определяется автоматически, ориентация – Горизонтальная. У нас нет многострочных полей, поэтому опция Однострочное должна быть отмечена для всех полей. Опцию Одно слово можно отметить для тех полей, где не могут встречаться пробелы, например, для поля E-mail.**
2. **Определите регистр букв. Можно оставить значение Любой для всех текстовых полей.**
3. **Поле Ваши комментарии и пожелания – является полем для ручного ввода. Введенный в него текст может быть написан сливающимися буквами и не распознается автоматически. Поэтому для этого поля на закладке Распознавание следует отметить опцию Не распознавать значение поля, оно будет введено вручную. Тогда значение в это поле будет вводить оператор.**
4. **Остальные свойства для всех полей можно оставить по умолчанию.**

#### 4.3.4.3.2. 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 шаблон).

Вы можете допустить исправления для каких-либо меток: в случае ошибочного проставления метки, заполняющий может просто заштриховать 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.

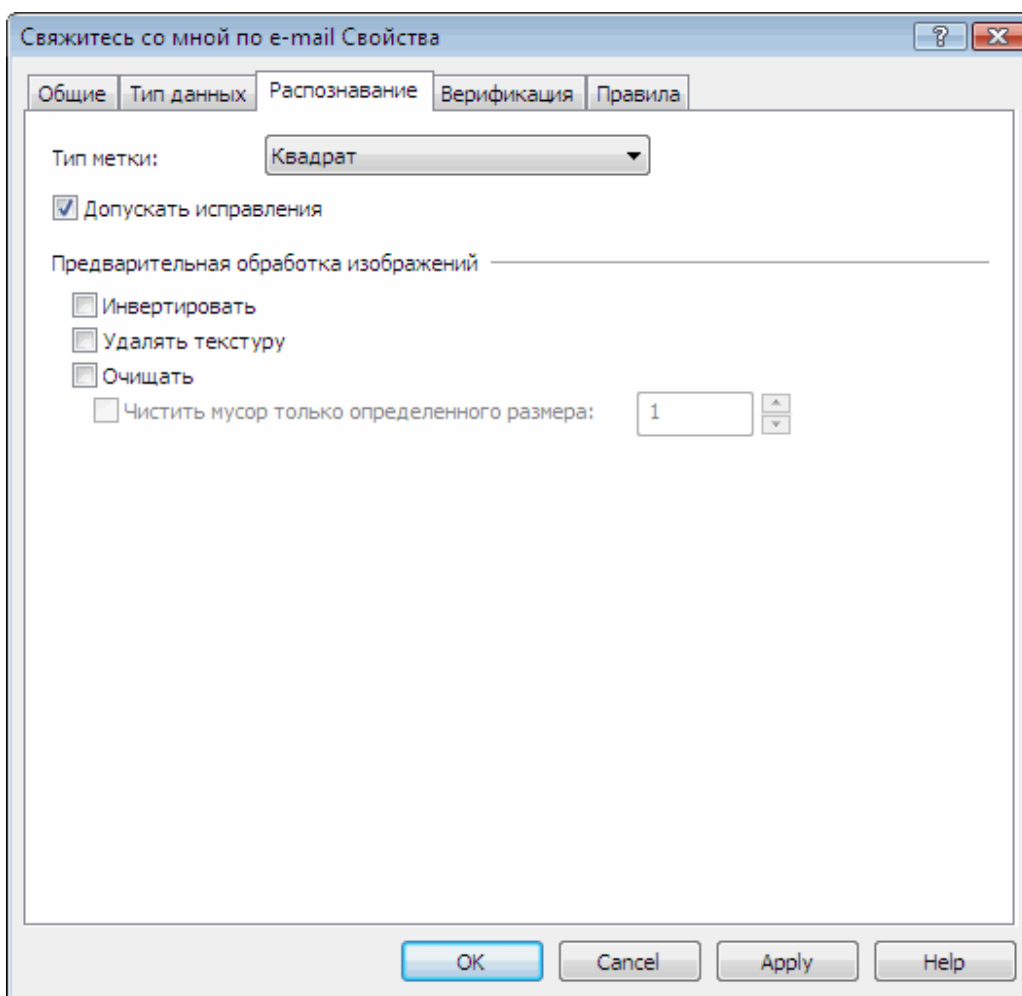


Рисунок 13. Диалог задания свойств поля, закладка для задания опций распознавания (для метки)

**Задайте свойства меток и групп меток.**

**Задайте тип *Квадрат*:**

- для меток, не входящих в группы;

- для групп меток.

**Можно отметить опцию *Допускать исправления*, допустив таким образом исправления для меток.**

#### 4.3.4.3.3. *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**.

#### 4.3.4.3.4. *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.)

#### 4.3.4.4. Параметры верификации

**Верификация** – это проверка распознанных данных оператором. В процессе создания шаблона вы можете задать опции верификации на закладке **Верификация** диалога задания свойств поля (Рисунок 14). 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.

Оператор видит распознанный символ и его изображение и либо соглашается с результатом распознавания, либо исправляет его.

Если вы хотите, чтобы все символы поля были включены в групповую верификацию, отметьте опцию **Включить в групповую верификацию**, а для того, чтобы поля участвовали в контекстной верификации – опцию **Включить в контекстную верификацию**. В групповой верификации могут участвовать все поля, но наиболее целесообразно отправлять на групповую верификацию метки и текстовые поля, содержащие только цифры с разделителями (точки, запятые, тире), а буквы лучше проверять в контексте значения поля. (Подробнее о групповой и контекстной верификации см. ниже в разделе Верификация).

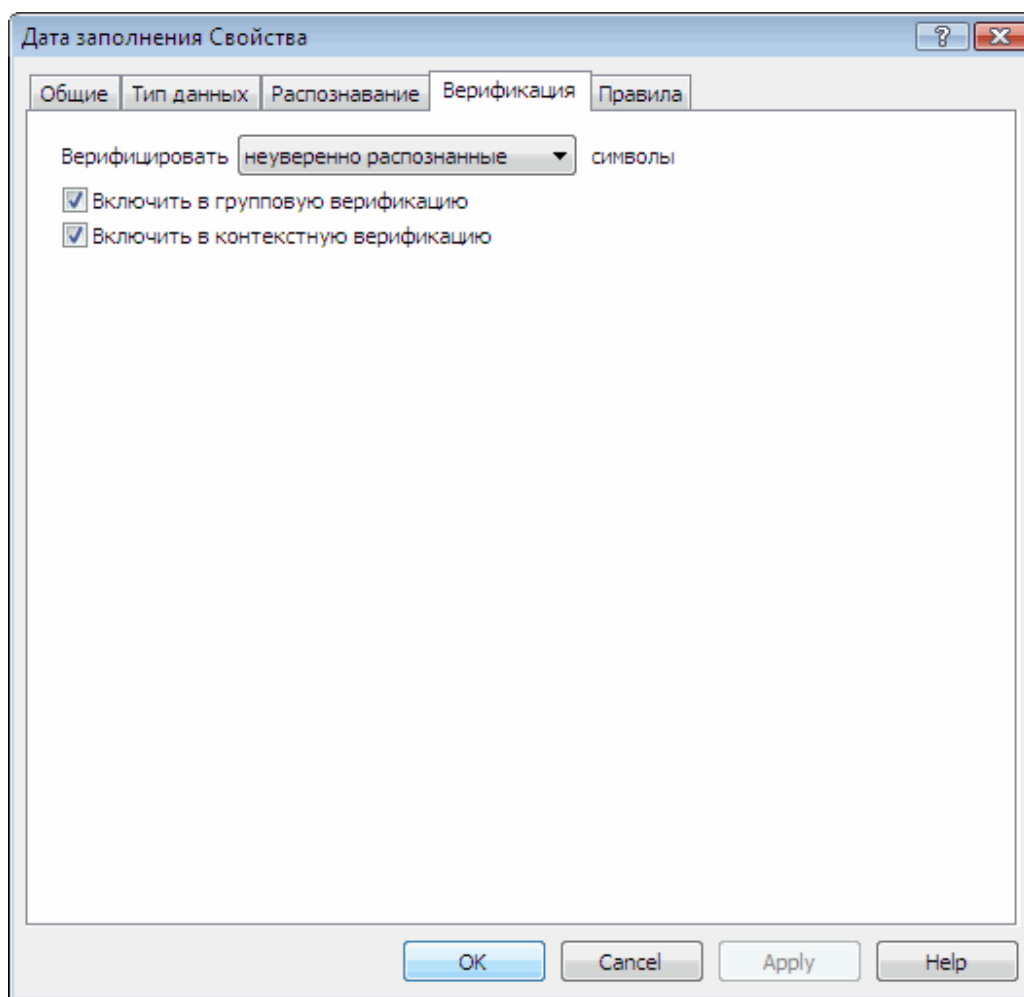


Рисунок 14. Диалог задания свойств поля, закладка для задания опций верификации

1. **Задайте опции верификации для полей ввода. По умолчанию выбрано значение *Верифицировать неуверенно распознанные символы*, то есть на верификацию будут подаваться неуверенно распознанные символы данного поля. Оставьте это значение для всех полей, кроме, например, имени, выбрав для этого поля значение *все*. Тогда все символы поля *Фамилия* и *Имя* будут подаваться на верификацию, вне зависимости от того, были ли они уверенно распознаны.**
2. **Отметьте опции *Включить в групповую верификацию* для всех меток и тех полей, которые будут заполнены цифрами (поля для ввода даты и количества страниц, обрабатываемых в день).**
3. **Включите опцию *Включить в контекстную верификацию* для полей, которые вы хотите проверить в процессе контекстной верификации (например, для имени, должности и названия компании).**

#### 4.3.4.5. Параметры экспорта картинок

Для поля Картинка в шаблоне 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.

#### 4.3.4.6. Проверка с помощью правил

**Правила** служат для автоматической проверки результата. 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 введенные значения: соединить значения нескольких полей или заменить распознанные значения соответствующими значениями из базы данных.

ABBYY FlexiCapture 8.0 Professional 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. 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 шаблона, 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 шаблона документа установлен русский язык распознавания.
- Скрипт – ограничения описываются пользователем с помощью скрипта.

Правила задаются на закладке **Правила** диалога задания свойств поля (Рисунок 15). 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.

При попытке экспорта документов с ошибками правил будет выводиться соответствующее предупреждение.

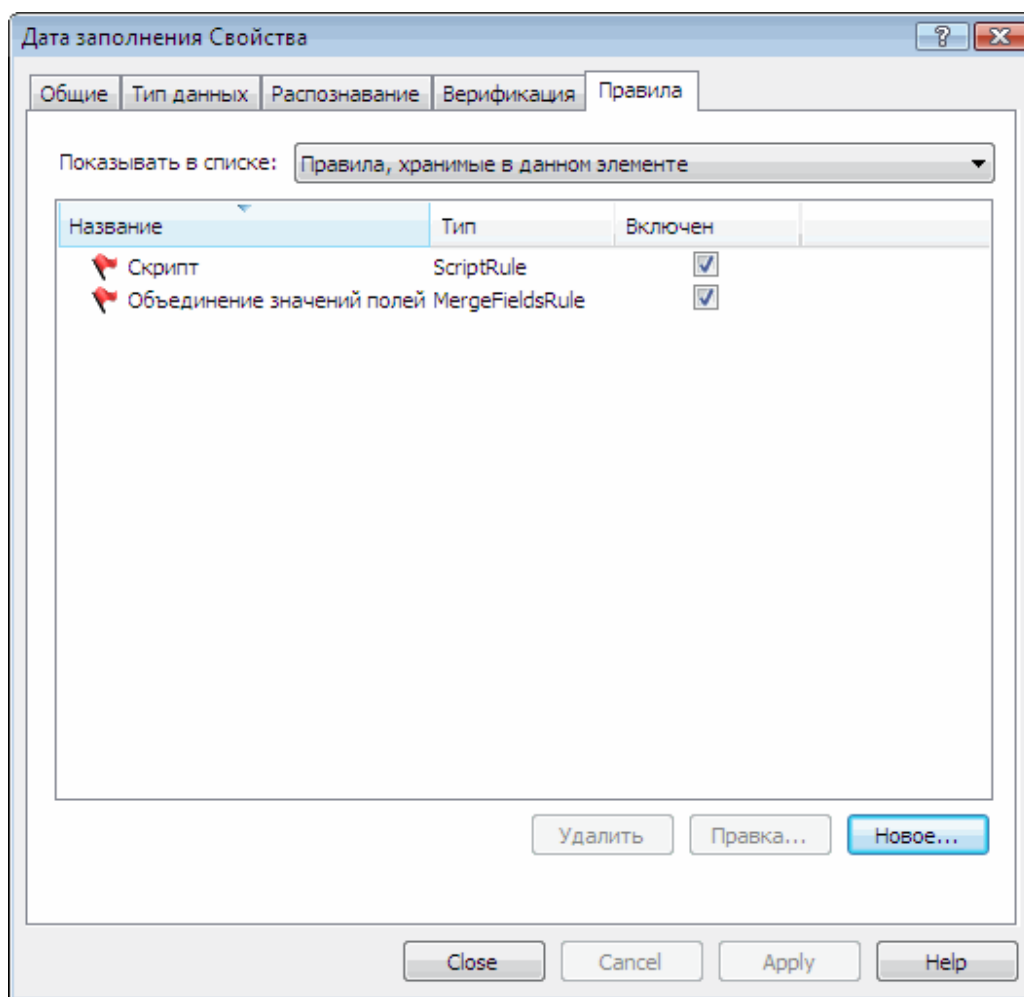


Рисунок 15. Диалог задания свойств поля, закладка для задания правил

Рассмотрим пример правила. Для этого представим дату заполнения анкеты в виде объединения трех полей: День, Месяц и Год. Для этого выполните следующие действия:

1. Удалите область поля даты на изображении, не удаляя само поле. Это можно сделать, выбрав в локальном меню поля *Дата заполнения* пункт *Удалить область*. Разметка исчезнет с изображения, но поле останется в структуре документа, помеченное красной звездочкой.
2. Создайте поля *День*, *Месяц* и *Год*.
3. Задайте свойства этих полей. На закладке *Общие* не отмечайте опцию *Экспортировать значение поля*, мы будем экспортировать только всю дату (проверьте, что опция *Экспортировать значение поля* отмечена у поля *Дата заполнения*).
4. На закладке *Тип данных* задайте тип данных для каждого из полей. Для каждого из них нужно выбрать тип данных *Число*, а формат – целое число. Кроме того, задайте ограничения для каждого поля, нажав кнопку *Правка...* в области *Проверка значения*:
  - День: от 1 до 31,
  - Месяц: от 1 до 12,
  - Год: в зависимости от периода заполнения анкеты, например, от 2000 до 2020.
5. На закладке *Распознавание* и *Верификация* задайте свойства, аналогично свойствам, которые мы задавали ранее для поля *Дата заполнения*.
6. Откройте диалог свойств поля *Дата заполнения* и перейдите на закладку *Правила*, чтобы приступить к созданию правила.
7. Нажмите на кнопку *Новое...* Выберите тип правила *Объединение значений полей*.
8. Присвойте правилу название, например, *Объединение значений полей – Дата заполнения*. Выберите *Ошибка*, чтобы присвоить правилу статус ошибки. Ошибка будет выдаваться в случае, если поле, участвующее в правиле, не будет обнаружено. Нажмите *Далее (Next)*.

9. Добавьте поля *День*, *Месяц* и *Год* в список полей для объединения значений, нажимая кнопку *Добавить...* В поле *Записать результат в поле* выберите поле *Дата заполнения*. В качестве разделителя используйте точку.

В поле *Дата заполнения* мы задавали проверку на попадание даты в указанный интервал. Теперь проверка будет осуществляться для значения, полученного путем слияния трех полей.

На этом заканчивается этап создания полей и статических элементов формы и задания их свойств. Проверьте, что вы создали все необходимые объекты и правильно задали их свойства. Если при наложении шаблона на изображение заполненной анкеты, некоторые элементы будут расположены неправильно или вас не удовлетворит результат распознавания, вы сможете вернуться к редактированию шаблона, изменить разметку или свойства некоторых полей.

#### 4.3.5. Как создать шаблон для многостраничных документов

ABBYY FlexiCapture 8.0 Professional позволяет создавать шаблоны для многостраничных документов. Шаблон может состоять из произвольного количества *разделов*, каждый из которых может содержать одну или несколько *страниц*. Для шаблонов многостраничных документов задается порядок следования разделов, их количество и правила сборки страниц в документы.

После завершения работы Мастера создания нового шаблона создается простой шаблон, не содержащий разделов. Вы можете добавлять в шаблон страницы. При добавлении страницы ее изображение добавляется в область изображения в окне редактора шаблонов, 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 шаблона документа.

Другой способ заключается в добавлении к шаблону разделов. В наиболее простом случае один раздел содержит одну страницу. Так можно поступать, если наборы полей на страницах независимы. Такой способ более нагляден: 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 – шаблон, состоящий из нескольких разделов, причем некоторые разделы содержат более одной страницы. Это, например, шаблон, состоящий из жестких разделов и гибкого многостраничного раздела. Или шаблон, описывающий документы, содержащие двусторонний лист, который может повторяться несколько раз. Допустим, вам нужно создать шаблон для 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.

Для шаблона, 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 шаблона документа **Шаблон > 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 **Шаблон > Добавить раздел документа...** в окне редактора шаблона. 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) либо 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 шаблона, в окне **Структура документа**, расположенном в правой части редактора шаблонов документов, 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.

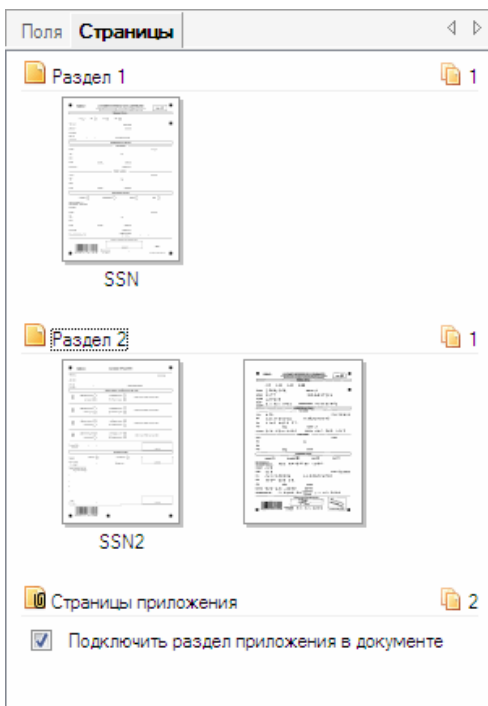


Рисунок 16. Структура документа, режим эскизов страниц

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

На закладке **Сборка документов** (Рисунок 17) 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.

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.

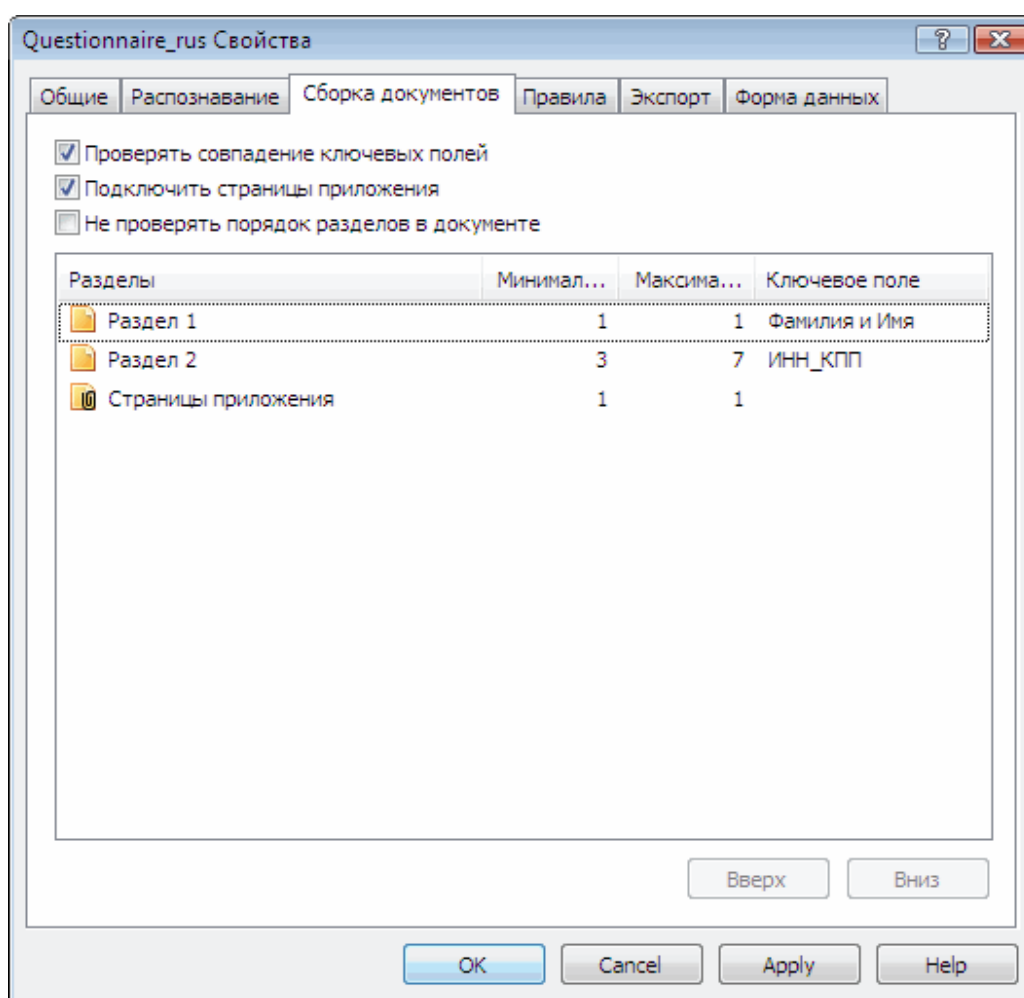


Рисунок 17. Диалог для задания структуры и правил сборки документов

Наша форма не является многостраничным документом, шаблон будет состоять из одной страницы, и правила сборки задавать не нужно. Вы можете проверить, что в диалоге свойств шаблона документа на закладке **Сборка документов** указан один раздел, повторяющийся один раз.

### 4.3.6. Как создать шаблон для документов с приложениями

ABBYY FlexiCapture 8.0 Professional позволяет создавать шаблоны для документов с приложениями.

Приложения – это дополнительные страницы, которые могут сопровождать каждый документ. С этих страниц не требуется извлекать отдельные поля, поэтому наложения шаблонов на них 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 шаблон документа с приложением:

- перейдите на закладку **Страницы** в окне **Структура документа** (правая часть редактора шаблона) и включите опцию **Подключить раздел приложения в документе**. Введите количество или диапазон страниц приложения в появившемся квадрате справа от названия раздела. (Рисунок 16)
- или откройте диалог свойств шаблона документа (**Шаблон > Свойства шаблона документа...** в окне редактора шаблона) и перейдите на закладку **Сборка документов**. Отметьте опцию **Подключить страницы приложения**. Введите минимальное и максимальное количество страниц в приложении (Рисунок 17).

Для того чтобы сохранить изображение приложения, нужно при настройке экспорта задать опции сохранения изображений (на закладке **Изображения** диалога **Настройки экспорта**). Изображение можно сохранить в каком-либо графическом формате, а также в PDF с возможностью поиска.

### 4.3.7. Настройка экспорта данных

Чтобы задать способ сохранения данных, полученных в ходе обработки бумажных документов, администратор настраивает экспорт для каждого шаблона документа. Возможен экспорт четырех типов: в файл заданного формата, в базу данных, в библиотеку документов MS SharePoint или экспорт, описанный с помощью скрипта. Для того чтобы настроить экспорт, в окне редактора шаблона документа выберите пункт меню **Шаблон > Настройки экспорта...** и задайте необходимые опции в диалоге **Настройки экспорта**.

Тип экспорта (файл, база данных, SharePoint или скрипт) нужно задать в поле **Вид экспорта**. От значения, выбранного в этом поле, зависит вид последующих настроек.

#### 4.3.7.1. Экспорт в файл

Если вы хотите настроить экспорт в файл, выберите значение **Экспорт в файлы** в поле **Вид экспорта** (Рисунок 18).

На закладке **Место сохранения** выберите папку для хранения файлов экспорта.

Вы можете экспортировать документы одного пакета в один файл (опция **Создавать папку для каждого пакета**) или каждый документ в отдельный файл (**Экспортировать каждый документ в отдельный файл**). Если вы не отметите ни одну из этих опций, все документы будут экспортироваться в один файл.

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.

На закладке **Формат** выберите тип файла (поддерживаются форматы CSV, DBF, TXT, XLS, XML) и задайте дополнительные свойства экспорта для выбранного типа файла. Вы также можете указать кодировку текста.

На закладке **Изображения** находятся параметры хранения изображений.

Отметьте опцию **Сохранять изображения документов**, чтобы сохранять изображения. Выберите директорию и имя файла, куда будут помещены обработанные изображения, либо отметьте опцию **в папку данных**, чтобы изображения сохранялись в той же папке, где будут сохранены данные.

Выберите формат хранения изображений. 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 шаблоне или выбрать из списка один или несколько языков (кнопка **Выбрать**).

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

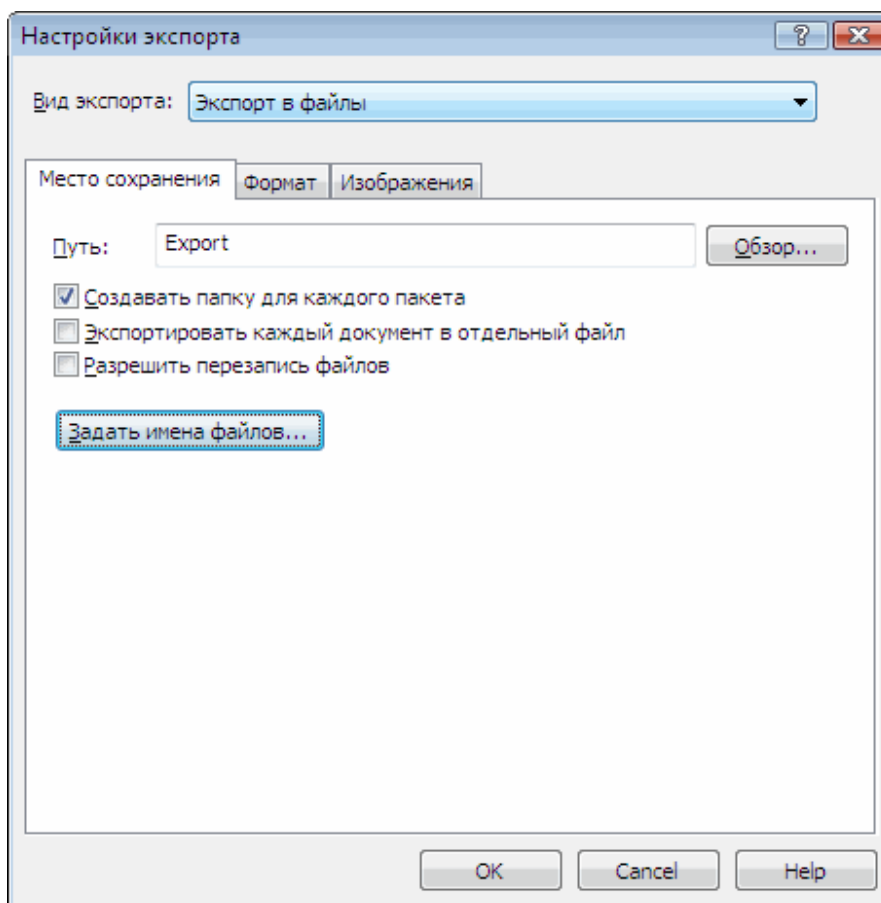


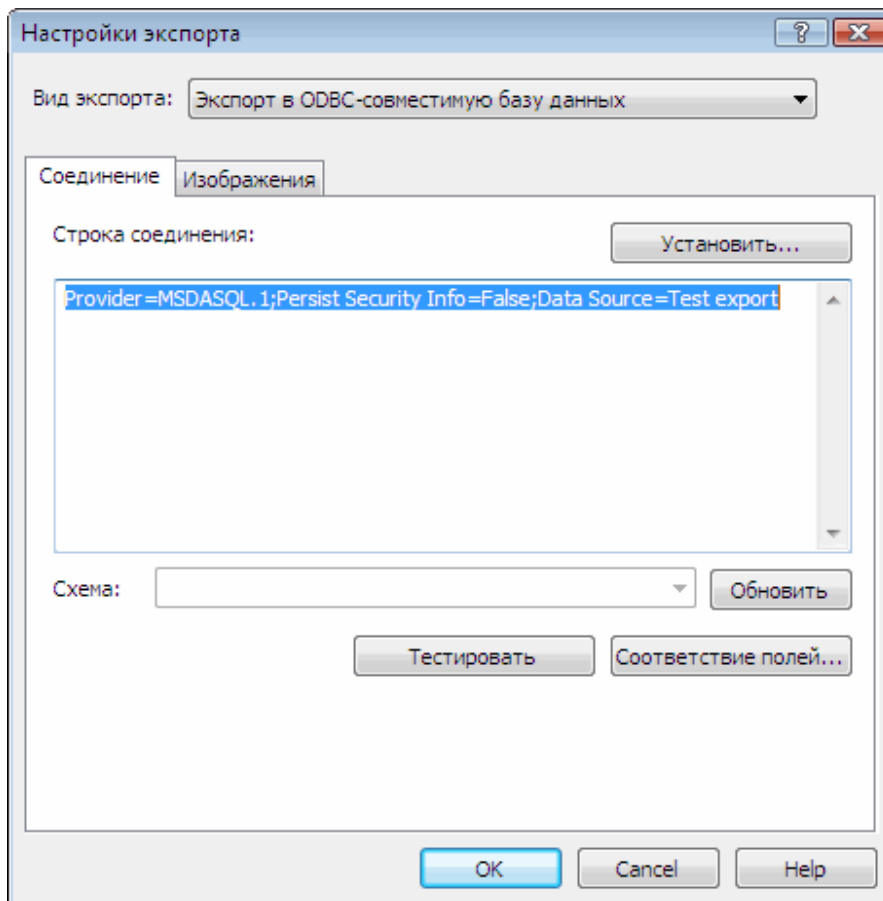
Рисунок 18. Диалог для задания настроек экспорта в файл

#### 4.3.7.2. Экспорт в базу данных

Для настройки экспорта в базу данных выберите значение **Экспорт в ODBC-совместимую базу данных** в поле **Вид экспорта** (Рисунок 19).

На закладке **Соединение** введите **Строку соединения** или нажмите на кнопку **Установить...** и задайте параметры в открывшемся диалоге (**Data Link Properties**). Выберите схему из выпадающего списка.

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



**Рисунок 19.** Диалог для задания настроек экспорта в базу данных

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

В диалоге **Соответствие полей** (Рисунок 20) 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.

Если в базе данных нет таблиц для экспорта документов, можно создать таблицы базы данных автоматически, для этого надо нажать на кнопку **Создать таблицы автоматически**. При ее нажатии создаются таблицы именно той структуры, которые нам 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 какое поле ключ должен быть проэкспортирован.

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.

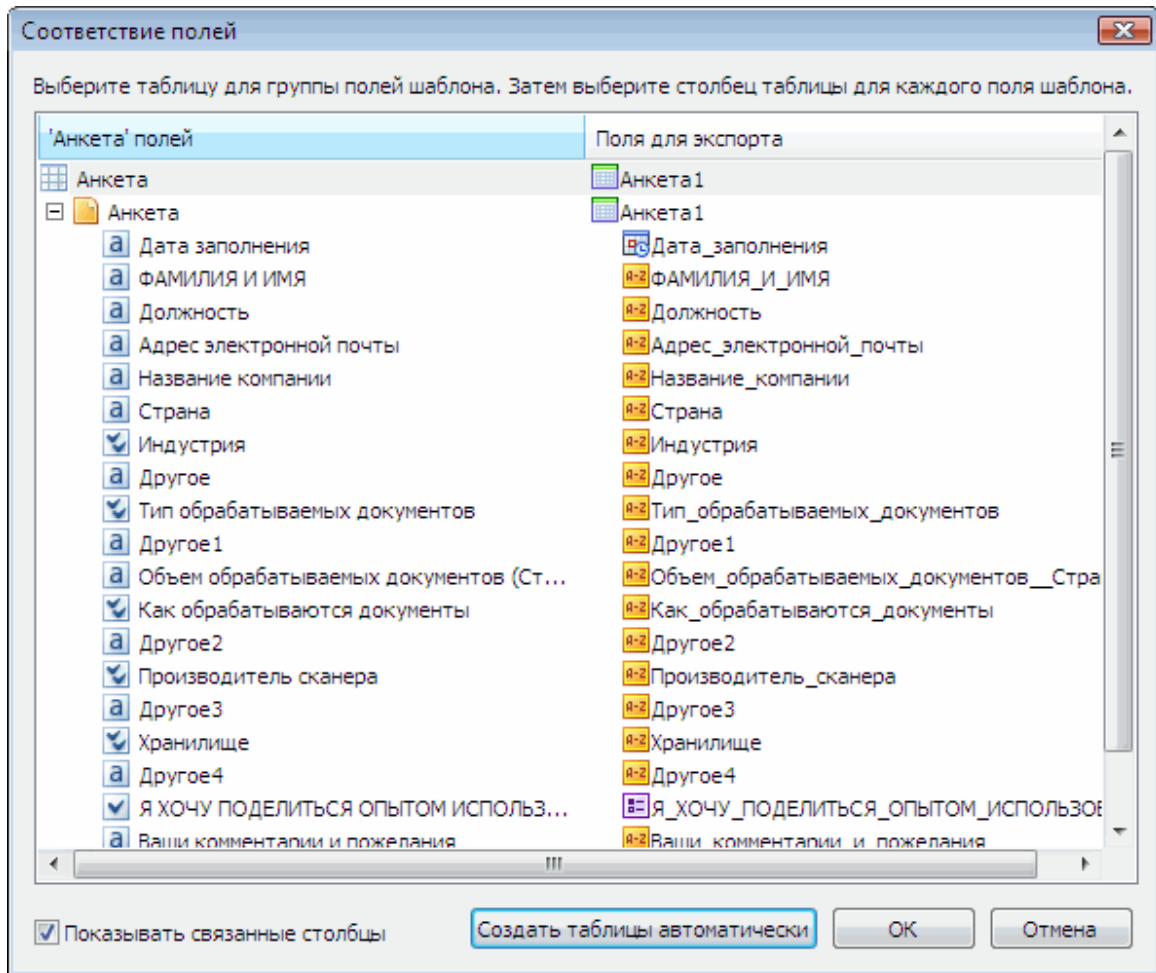


Рисунок 20. Задание соответствия полей документа полям базы данных при экспорте

На закладке **Изображения** диалога **Настройки экспорта** можно задать параметры хранения изображений.

Вы можете сохранять изображения в базу данных или в файл (в этом случае укажите папку для сохранения).

Выберите один из форматов хранения изображений. Если вы выберете формат PDF с возможностью поиска, то будет сохранено не просто изображение документа, а будет произведено его полнотекстовое распознавание, а распознанный текст сохранится в выбранном формате.

Если вы хотите изменять разрешение исходных изображений, отметьте опцию **Изменить разрешение и установить его равным** и введите разрешение.

#### 4.3.7.3. Экспорт в библиотеку документов SharePoint

Программа ABBYY FlexiCapture 8.0 позволяет настроить экспорт документов в библиотеку Microsoft™ SharePont. 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 **Вид экспорта** (Рисунок 21).

На закладке **Соединение с SharePoint** введите адрес сервера (**URL сервера**), где находятся ваши 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**.

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

На закладке **Изображения** нужно задать параметры хранения изображений (формат файла, качество изображения и тип цвета, установите разрешение). Если вы выбрали формат PDF или PDF/A для хранения изображений, то вы также можете создать PDF с возможностью поиска, для этого следует отметить опцию **Создать PDF с возможностью поиска**. Вы можете выбрать язык (или несколько языков, если ваш документ написан на нескольких языках), отличный от языка, заданного в шаблоне документов. Для этого нажмите на кнопку **Выбрать...** справа от поля **Язык** и добавьте нужные языки. Если вы хотите использовать язык шаблона, то в поле **Язык** должно быть значение **Как в шаблоне**.

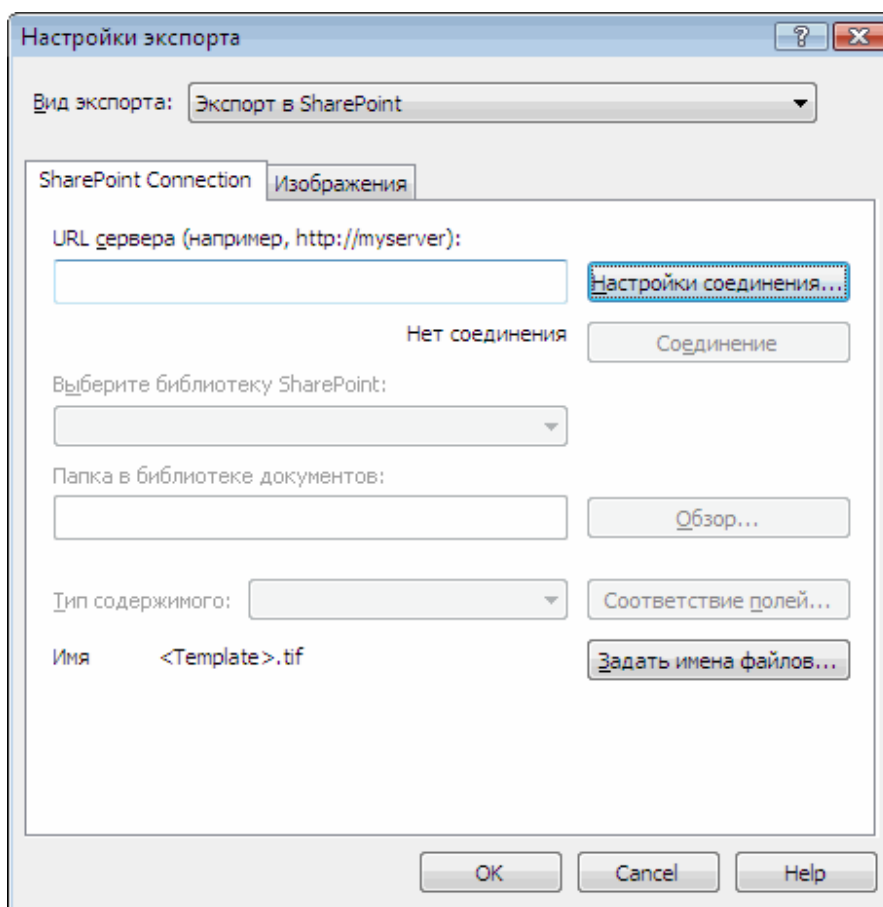


Рисунок 21. Диалог для задания настроек экспорта в библиотеку документов SharePoint

#### 4.3.7.4. Экспорт, описанный с помощью скриптовой процедуры

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

Если вы хотите настроить экспорт с помощью скриптовой процедуры, выберите значение **Пользовательский экспорт (скрипт)** в поле **Вид экспорта** (Рисунок 22).

Затем выберите язык скрипта (**JScript®** или **VBScript**) и введите текст скрипта в окне редактора, которое появляется при нажатии кнопки **Редактировать...** (Подробное описание скриптов с примерами использования вы найдете в справке программы.)

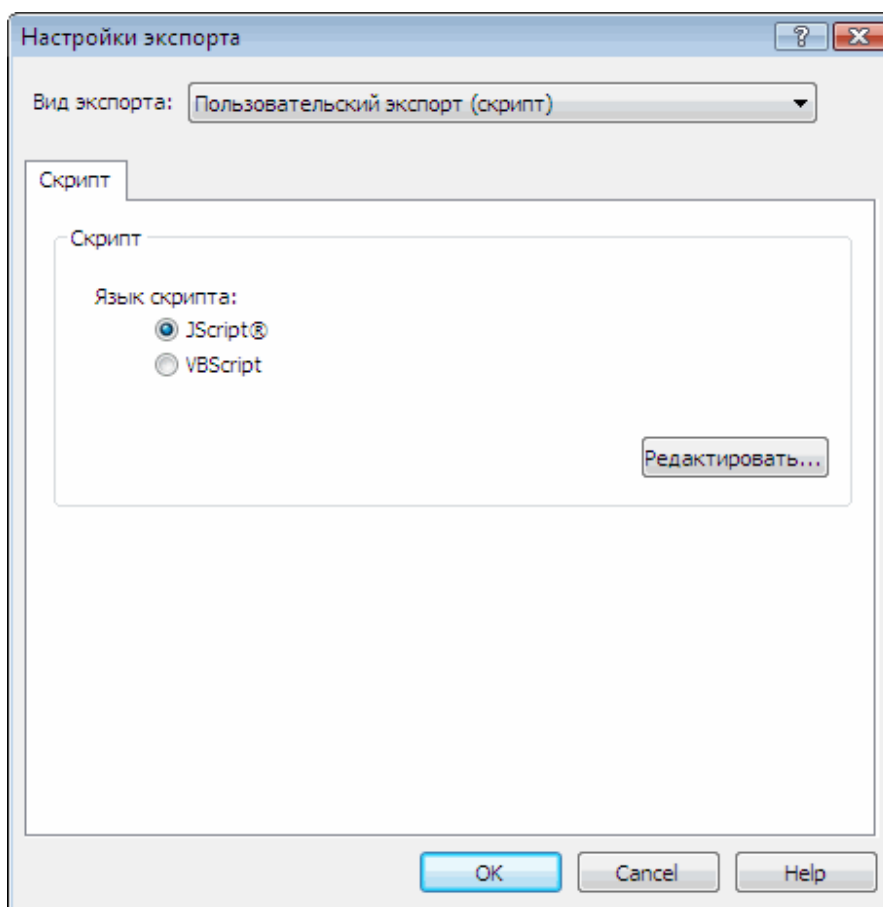


Рисунок 22. Диалог для задания настроек экспорта с помощью скрипта

1. В окне редактирования шаблона выберите пункт меню *Шаблон>Настройки экспорта*. В поле *Вид экспорта* открывшегося диалогового окна выберите значение *Экспорт в файлы*.
2. На закладке *Место сохранения* введите путь к директории, в которой будет сохранен файл экспорта, и задайте принцип формирования имен файлов.
3. На закладке *Формат* выберите тип файла *XML Document (\*.xml)* и отметьте опцию *Ошибки* в секции *Экспортировать дополнительные данные*.
4. Если вы хотите сохранять изображения страниц, то перейдите на закладку *Изображения*, отметьте опцию *Сохранять изображения документов* и задайте необходимые настройки.
5. После задания настроек экспорта, основные параметры будут отображены на закладке *Экспорт* диалогового окна свойств шаблона документа (*Шаблон>Свойства шаблона документа...*). Вы также можете изменять параметры экспорта из этого диалогового окна, нажав кнопку *Изменить настройки...*

#### 4.3.8. **Настройка отображения распознанных данных**

После того, как данные будут распознаны, пользователь увидит их в окне документа. По умолчанию данные будут расположены по порядку, подписи к ним будут соответствовать названиям полей шаблона. 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 шаблона 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 шаблона документа (Шаблон > Свойства шаблона документа...)**.

Посмотрите, как будут располагаться распознанные данные в окне документа. Вы можете оставить расположение данных по умолчанию либо изменить его.

На этом заканчивается создание шаблона.

#### **4.3.9. Редактирование и публикация шаблона**

После создания шаблона документа его нужно сохранить и протестировать на нескольких изображениях. Получив удовлетворительный результат, шаблон нужно опубликовать, чтобы он стал доступен для распознавания документов.

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

Чтобы вернуться к редактированию шаблона, выделите его в диалоге **Шаблоны документов** и нажмите на кнопку **Правка...** Шаблон будет недоступен для редактирования другими пользователями, а распознавание будет производиться с использованием последней опубликованной версии, так что редактирование шаблона не помешает работе операторов. После редактирования опубликуйте шаблон. Всем 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 **Шаблоны документов** нужно выделить шаблон и нажать на кнопку **Отменить правку**.

#### **4.3.10. Тестирование шаблона**

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

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

Дальнейшее тестирование производится на документах, добавляемых в тестовые пакеты, т.к. эти пакеты работают с локальной версией шаблона. Эти пакеты доступны из меню **Проект>Тестовые пакеты** главного окна программы. Также к ним можно перейти из окна редактора шаблона документа, выбрав пункт меню **Сервис>Тестовые пакеты**.

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

Протестируйте шаблон документа сначала на исходном изображении (пункт меню **Документ>Начать тестирование**), а затем на заполненном изображении анкеты (см. **Работа оператора с настроенным проектом**) и при необходимости внесите какие-либо изменения. Опубликуйте шаблон.

## 4.4. Настройка импорта изображений

Оператор начинает работу с добавления в проект новых изображений: это могут быть бумажные документы, которые он должен отсканировать, или уже имеющиеся электронные изображения. Часто выполняемые действия по добавлению регулярно поступающих из одного и того же места изображений можно автоматизировать, тогда импорт будет производиться нажатием одной кнопки. Если произведены необходимые настройки, то при нажатии одной кнопки выполняется сканирование или импорт из заранее настроенного источника, а полученные изображения обрабатываются (например, очищаются от «мусора»). При желании оператор может изменить источник импорта изображений или его настройки.

В случае если источники импорта часто меняются, то во избежание необходимости настройки источников можно создать соответствующие им профили и переключаться между ними.

Кроме того, можно настроить периодическое добавление изображений в фоновом режиме.

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 23).

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.

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 сборки страниц в 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 удалять или перемещать в другую папку.
5. На последнем шаге мастера вы можете изменить имя, присвоенное профилю импорта по умолчанию, и ввести комментарий.

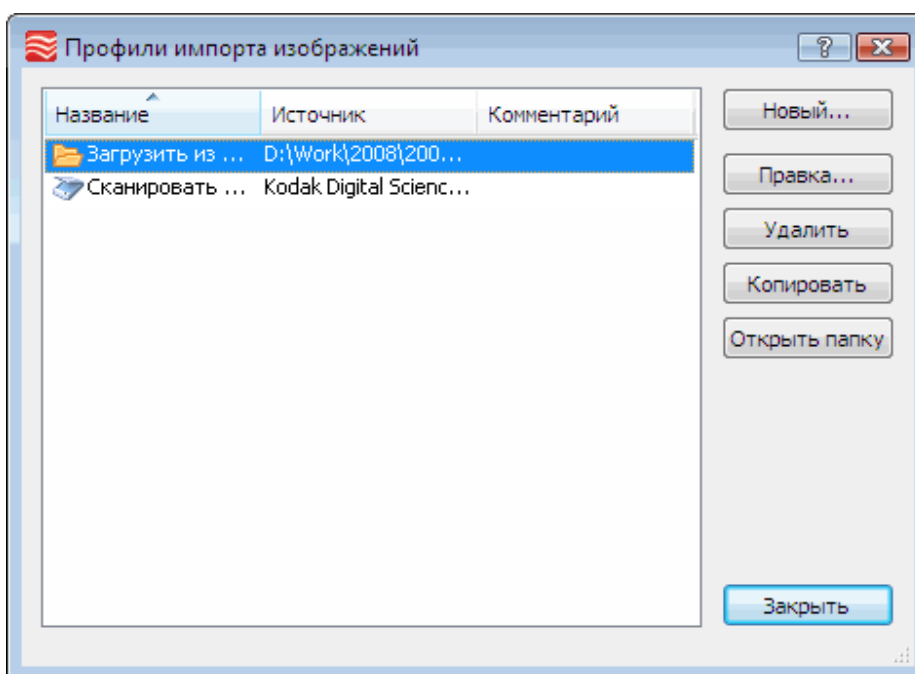


Рисунок 23. Диалог создания и редактирования профилей импорта

1. Настройте профиль импорта для ваших изображений. Для этого перейдите в главное окно программы ABBYY FlexiCapture 8.0 Professional и выберите пункт меню *Проект > Профили импорта изображений*.
2. В диалог *Профили импорта изображений* нажмите *Новый...* для создания нового профиля, введите название профиля, выберите опцию *Сканер*, так как изображение заполненной формы поступит со сканера
3. Задайте опции сканирования и обработки поступающих изображений.
4. На последнем шаге мастера вы можете изменить имя, присвоенное профилю импорта по умолчанию, и ввести комментарий.

## 5. 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 шаблон с жестким расположением полей, потому что местоположение, размеры и количество полей на разных экземплярах могут не совпадать, а значит, жесткий шаблон не может быть наложен на все документы. На таких документах поля ищутся после предварительного распознавания всей страницы целиком, а шаблон создается на основе ключевых слов и взаимного расположения полей. Для разработки таких шаблонов предназначен специальный компонент программы АBBYY FlexiLayout Studio 8.0. Подробное руководство создания гибкого шаблона вы найдете в справке для АBBYY FlexiLayout Studio 8.0.

Созданный гибкий шаблон экспортируется в .afl файл, а затем подключается в АBBYY FlexiCapture 8.0 Professional. Мы можем использовать гибкое описание для всего шаблона или в качестве одного из разделов. Гибкое описание может быть многостраничным.

Гибкое описание можно подключить на этапе создания шаблона документов. Для этого на втором шаге Мастера создания шаблона 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 шаблонов документов, в диалоге свойств раздела шаблона 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 шаблоне 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 шаблона. Определите свойства распознавания, верификации, экспорта, задайте необходимые правила.

Подробнее см. [Создание шаблона документа](#).

## 6. 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 АBBYY FlexiCapture 8.0 Professional. 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 шаблон документов 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.

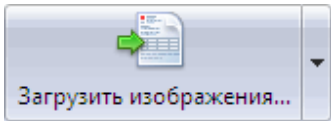
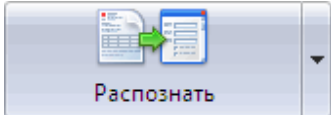
Для хранения документов нужно настроить экспорт: значения ключевых полей вы можете экспортировать в файл или базу данных, а изображения документов сохранить в удобном для вас формате. Для этого на закладке **Изображения** диалога **Настройки экспорта** отметьте опцию **Сохранять изображения документов**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

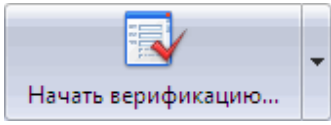
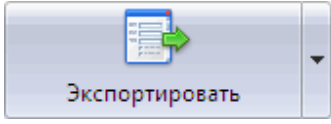
Обратите внимание на сборку страниц в документы: для неструктурированных документов бывает сложно определить, к какому документу относится та или иная страница. Для автоматизации сборки неструктурированных документов рекомендуется разделять документы чистыми страницами или страницами со штрих-кодами. 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.

## 7. Работа оператора с настроенным проектом

После того как администратор разработал шаблон и произвел все необходимые настройки, можно приступать к обработке документов. Весь процесс можно настроить таким образом, чтобы операции добавления изображений, распознавания и экспорта данных выполнялись с минимальным участием человека. Работа оператора не зависит от вида документов.

Процесс обработки документов состоит из четырех этапов и каждому этапу соответствует кнопка на панели инструментов.

<p><b>1. Импорт изображений</b></p> <p>Стрелка справа позволяет выбрать одну из опций импорта изображений: <b>Загрузить изображения...</b>, <b>Сканировать изображения</b> или произвести импорт, выбрав один из настроенных профилей. Надпись на кнопке соответствует выполняемому действию.</p>	
<p><b>2. Распознавание</b></p> <p>Запускает процесс распознавания документов. Нажав на стрелку справа, вы</p>	

также можете произвести анализ документов или наложение одного из шаблонов.	
<b>3. Верификация</b> Запускает процесс верификации. Нажав на стрелку справа, можно также запустить процесс проверки правил.	
<b>4. Экспорт</b> Запускает экспорт данных согласно определенным в шаблоне настройкам. Нажав на стрелку справа, можно осуществить экспорт в файл или базу данных.	

## 7.1. Добавление изображений

Обработка документов начинается с добавления в проект изображений страниц. Для начала работы выберите или создайте новый пакет, в который будут добавлены изображения (для создания щелкните правой кнопкой мыши в главном окне программы и выберите **Новый пакет**). При попытке добавить изображения в проект, где нет ни одного пакета, пакет документов будет создан автоматически.

Добавить изображения в пакет можно несколькими способами:

1. Загрузить ранее сохраненные изображения из файлов. Для того чтобы загрузить изображения из файла, выберите **Загрузить изображения...** в выпадающем меню кнопки импорта или нажмите комбинацию клавиш **Ctrl+O**. В открывшемся диалоге выберите нужные файлы с изображениями и задайте опции импорта изображений.

При импорте изображений из многостраничных файлов, в пакет будут добавлены несколько страниц.

2. Отсканировать изображения. Для добавления изображений со сканера, выберите пункт меню **Сканировать изображения...** Вам будет предложено выбрать сканер и отсканировать изображения.
3. Произвести импорт изображений, используя один из ранее настроенных администратором профилей импорта изображений (см. [Настройка импорта изображений](#).)

Если профили импорта были настроены, то их названия отобразятся в меню кнопки импорта. Для начала импорта изображений достаточно выбрать пункт с одним из названий профилей импорта.

При выборе **Импортировать изображения...** откроется диалог **Выбор профиля импорта**, в котором нужно выбрать профиль импорта из выпадающего списка и нажать кнопку **Импортировать** для начала импорта изображений. Из этого диалога можно также создать новый профиль импорта или отредактировать существующие профили.

Выбрав один раз профиль импорта, для повторного использования того же профиля уже не нужно искать его в списке, достаточно нажать на кнопку импорта, на которой отображается название последнего используемого профиля.

Импорт изображений может производиться в фоновом режиме, если настроен соответствующий профиль импорта. Изображения из горячей папки автоматически импортируются в систему, проверка наличия новых изображений производится через указанный в настройках интервал времени.

После того, как изображения добавлены в пакет, необработанные страницы появляются в списке.

1. **Создайте пакет, в который будет добавлено изображение. Для этого в главном окне программы щелкните правой кнопкой мыши на белом фоне и выберите *Новый пакет*. Введите название пакета и комментарий. Откройте пакет двойным щелчком мыши.**
2. **Заполните анкету.**
3. **Отсканируйте изображение.**
4. **Выполните импорт изображения, выбрав *Импортировать изображения...* или *Сканировать изображения...* Страница с названием <Неизвестный> появится в пакете.**

## 7.2. Распознавание

Распознавание производится при нажатии кнопки **Распознать**. При этом программа производит наложение шаблона, выбирая его по реперам и идентификаторам, и производит распознавание данных в областях документов, заданных полями выбранного шаблона.

Распознавание изображений может производиться автоматически, сразу после добавления изображений в пакет. Для этого нужно отметить опцию **Автоматически распознавать добавленные изображения** (меню **Сервис** главного окна, пункт **Опции...**, закладка **Обработка документов**).

Процентное соотношение уверенно распознанных символов к общему количеству символов отображается в колонке **Уверенность распознавания**. После распознавания оператор может перейти к верификации документов.

Еще до начала распознавания страницы можно *проанализировать*, то есть наложить шаблон на страницу. Для этого нужно выбрать пункт **Проанализировать** выпадающего меню кнопки распознавания. Если шаблон может быть наложен, то у страницы и документа, в который входит страница, появляются соответствующие названия и происходит наложение полей. Если ни один из шаблонов проекта не может быть наложен на изображение, то страница остается необработанной.

В большинстве случаев правильно созданный в программе шаблон накладывается на страницы автоматически, однако в некоторых случаях может возникнуть необходимость выбрать шаблон для наложения вручную. Для наложения шаблона выделите документ или страницу и выберите пункт **Наложить шаблон...** выпадающего меню кнопки распознавания.

**Выберите команду *Проанализировать* выпадающего меню кнопки распознавания. Если шаблон успешно наложен, то его название отобразится вместо <Неизвестный>.**

**Нажмите кнопку *Распознать*. Страница будет распознана.**

## 7.3. Верификация

Верификация, или проверка распознанных данных – самая трудоемкая часть работы оператора. Поэтому процесс верификации в ABBYY FlexiCapture 8.0 Professional организован таким образом, чтобы быть максимально удобным, и чтобы количество ошибок сводилось к минимуму. Для многостраничных документов на первом этапе производится проверка правильности сборки страниц в документы. Далее, запускается процесс верификации распознанных данных, который

состоит из групповой и контекстной верификации. Можно также произвести проверку в окне документа. В процессе верификации производится проверка правил.

**Проверка сборки документов.** Для многостраничных документов производится проверка структуры документа. Если порядок страниц документа не совпадает с заданным или не совпадают значения ключевого поля на страницах документа, то такой документ помечается красным флажком, а в окне документа отображается сообщение об ошибке. Оператор должен проверить, не были ли перепутаны страницы при сканировании. Иногда ошибки сборки можно исправить, просто изменив порядок страниц.

Сборку страниц в документы удобно проверять при использовании режима эскизов страниц (Рисунок 24). Вы можете изменять расположение страниц и даже переносить их из одного документа в другой, просто перетаскивая их с помощью мыши.

Если в шаблоне задано, что для правильной сборки страниц должна производиться проверка по ключевым полям, то значения ключевых полей будут отображены под изображением каждой страницы. В случае несовпадения ключевых полей на страницах одного документа, они будут отображены красным цветом. Ключевые поля могут не совпадать из-за того, что они были неправильно распознаны или заполнены. Проверьте значения ключевых полей. Если же значения ключевых полей все равно не совпадают, то, вероятно, такие страницы не относятся к одному документу. Если был перепутан порядок страниц, то найдите страницы с одинаковыми ключевыми полями и соберите их в документы.

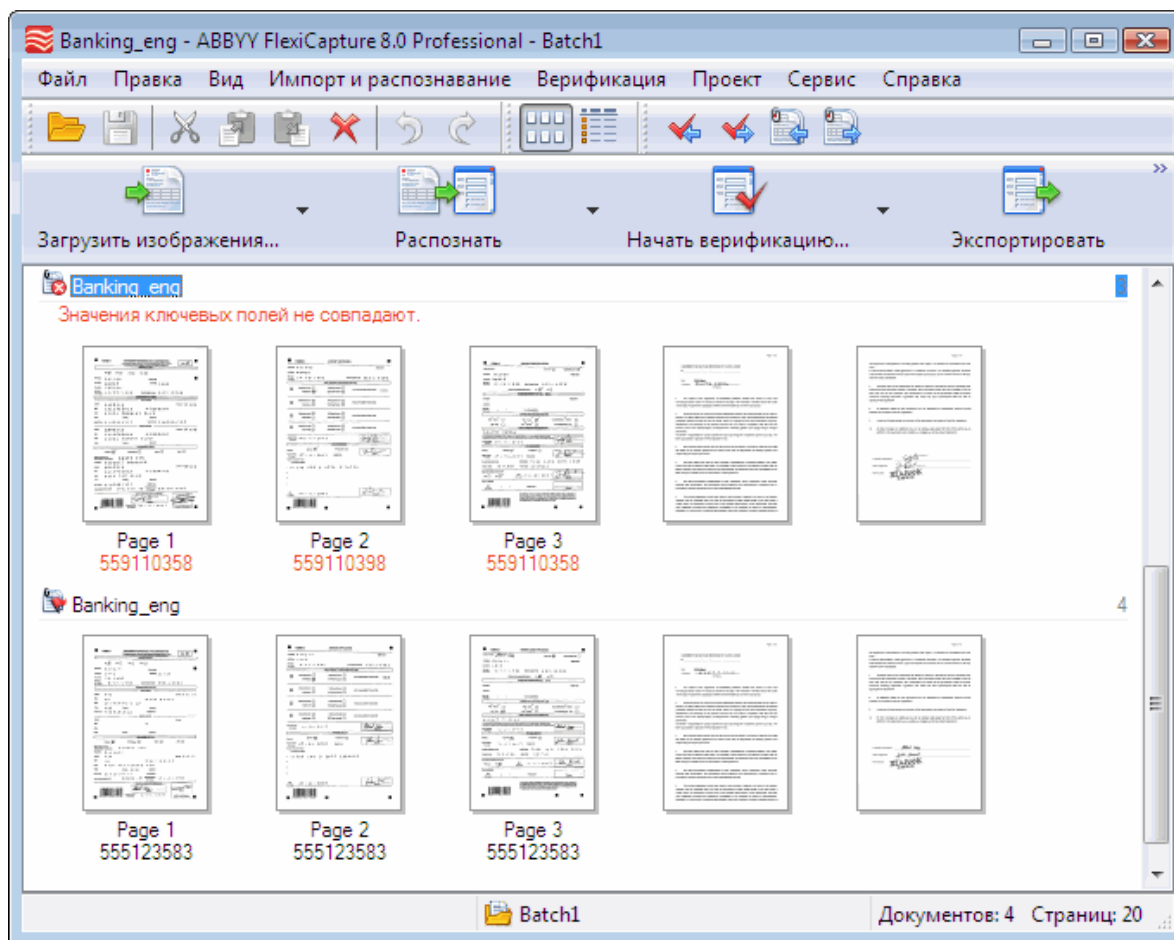


Рисунок 24. Главное окно в режиме просмотра эскизов страниц

**Замечание.** Вы можете изменять масштаб эскизов страниц для просмотра деталей, для этого нажмите Ctrl и вращайте колесо мыши.

Процесс верификации распознанных данных запускается при нажатии кнопки **Начать верификацию...**

**Групповая верификация** – это группировка изображений символов, распознанных с одинаковым значением, и их вывод на экран верификатора с тем, чтобы подтвердить явно верные варианты, и оставить на следующий этап только те, которые либо неверны, либо сомнительны (Рисунок 25).

При групповой верификации оператор может просмотреть изображение поля, в котором встретился проверяемый символ, выбрав пункт **Показать изображение символа** в локальном меню выделенного символа или нажав F2. Кроме того, можно включить режим, при котором будет отображаться поле, содержащее символ, при наведении курсора мыши на проверяемый символ. Для этого нужно выбрать пункт меню **Вид > Изображение поля > Показывать изображение поля** в окне верификатора, либо нажать комбинацию клавиш Ctrl+I.

Неправильно распознанные символы можно исправить следующим образом: выделите символ, который не соответствует символу группы, и введите его правильное значение. Это значение отобразится зеленым цветом в левом верхнем углу изображения символа. Если вы сомневаетесь в значении символа даже после просмотра контекста, нажмите на символ левой кнопкой мыши: символ будет помечен красным знаком вопроса. Статус символа можно изменять также с помощью кнопки на панели инструментов **Изменить статус**.

Правильно распознанные символы можно подтвердить с помощью локального меню, выбрав пункт **Подтвердить**. Можно также подтвердить сразу все отображаемые символы, нажав кнопку **Подтвердить все** на панели инструментов.

В процессе создания шаблона документов, опции верификации задаются при определении свойств поля. На групповую верификацию подаются символы из тех полей, для которых отмечена опция **Включить в групповую верификацию** на закладке **Верификация** в диалоге задания свойств поля.

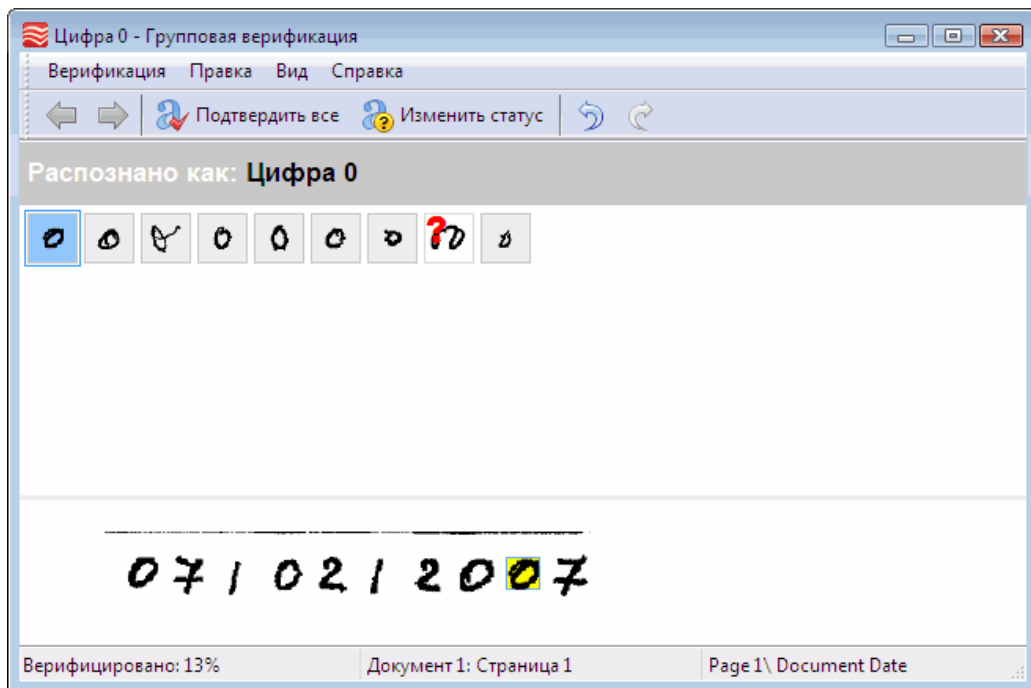


Рисунок 25. Групповая верификация цифр

© 2009 ABBYY. Все права защищены.

**Контекстная верификация** – это режим верификации, предназначенный для эффективного исправления ошибок формата полей, область значений которых известна или легко идентифицируется. Таким полем, например, является название страны: нам известно, какие значения может принимать данное поле.

Для принятия решения о том, как нужно исправить поле, пользователю достаточно видеть только это поле и знать его область значений. Исправляются не только ошибки заполнения, но и ошибки формата поля.

На контекстную верификацию подаются те поля, для которых отмечена опция **Включить в контекстную верификацию** на закладке **Верификация** в диалоге задания свойств поля.

Для исправления неправильно распознанных символов можно использовать стандартные режимы текстовых редакторов: режим вставки и замены символов, переключаться между ними можно с помощью кнопки **Insert** на клавиатуре.

Пользователь последовательно просматривает результат распознавания каждого поля, исправляет его и подтверждает, нажав **Enter** или кнопку **Подтвердить поле**.

Поля, значения которых не соответствуют заданному типу данных, помечаются красным флажком, для них выводится сообщение об ошибке. Так же помечаются поля, для которых были заданы правила, и значение поля не удовлетворяет условию. Значения в таких полях нужно исправить, а если исправить не удастся, то отложить проверку значения поля с помощью кнопки **Отложить поле**, чтобы внести исправления позже.

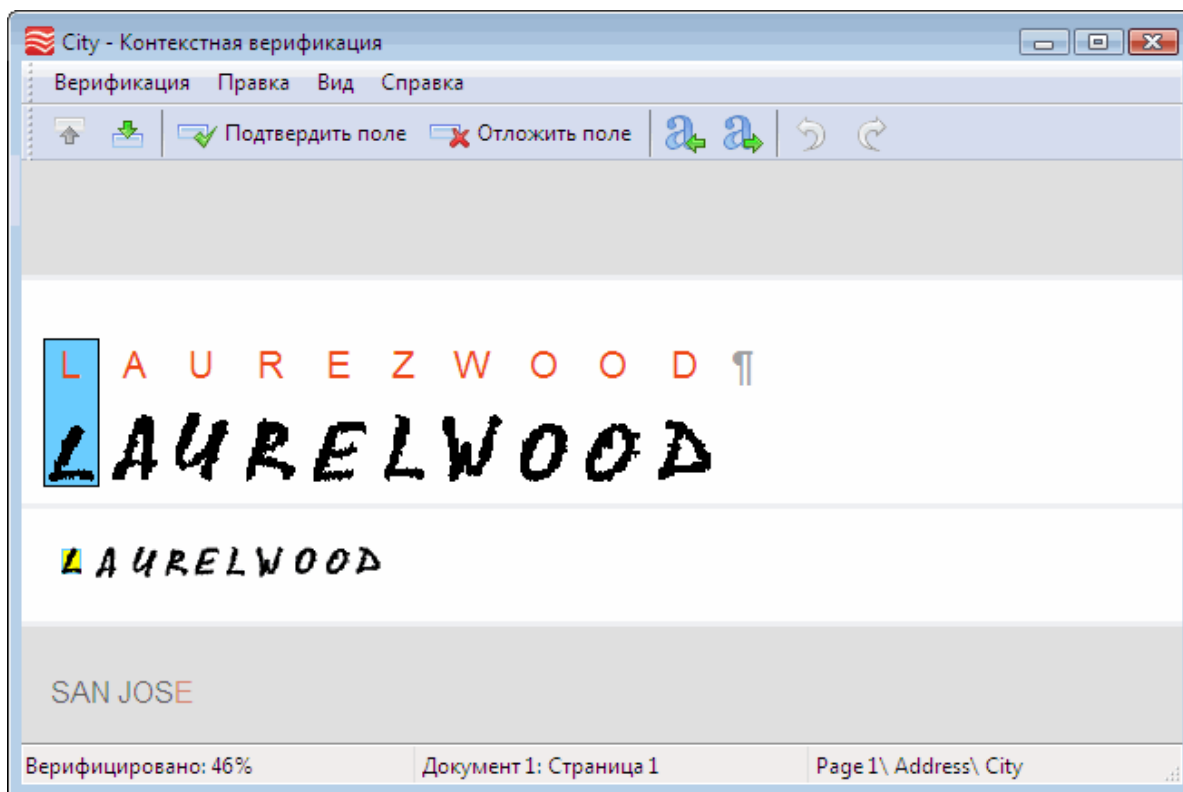




Рисунок 26. Окно контекстной верификации

**Окно документа** также позволяет проверить правильность распознавания и исправить неверно распознанные символы (Рисунок 27). Окно документа открывается при двойном щелчке на названии страницы и состоит из области данных, изображения страницы и области ошибок правил (если такие ошибки есть). Удобное расположение окон можно настроить с помощью кнопки **Вид**

**окна.** В окне документа оператор видит уже не группы символов или отдельные поля, а весь документ целиком.

В области данных неуверенно распознанные и сомнительные символы выделяются красным цветом, поля с неверными типами данных и невыполненными правилами подсвечиваются. Для быстрого перехода к предыдущей и следующей ошибке можно использовать кнопки  и . По этим кнопкам осуществляется последовательный переход по ошибкам сборки, неуверенно распознанным символам, ошибкам правил и т.д.

Поля для просмотра и редактирования распознанных данных могут располагаться не только в порядке следования, но и в любом другом порядке, удобном для пользователя. Их расположение можно изменить в редакторе шаблона документа, см. [Настройка отображения распознанных данных](#).

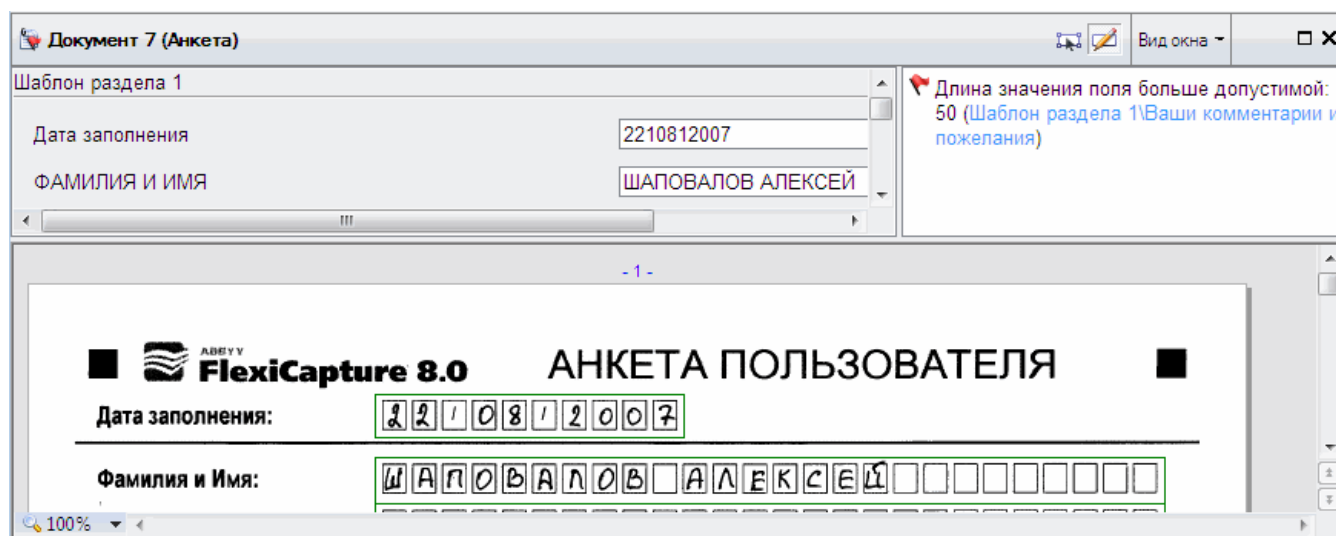


Рисунок 27. Окно документа

**Проверка правил.** Невыполненные правила помечаются флажком: желтым в случае предупреждения, красным в случае ошибки. Если правило относится к одному из полей, то такое поле подается на проверку оператору в процессе контекстной верификации. Ошибки правил отображаются в отдельном окне редактора документов, а документы с невыполненными правилами отмечены красными флажками.

При наличии невыполненных правил оператору надо проверить, как распознаны данные, если это ошибка распознавания – исправить ее, а если это ошибка заполнения, которую невозможно исправить, не экспортировать документ.

Для запуска повторной проверки ошибок правил нажмите на стрелку справа от кнопки и выберите пункт меню **Перепроверить правила**. Если правила выполняются после исправления значений полей, то флажки будут сняты.

1. Нажмите кнопку **Начать верификацию...** и произведите проверку данных.
2. Посмотрите, какие символы и поля подаются на групповую и контекстную верификацию. Проверьте их.
3. Если какие-либо поля помечены красным флажком с сообщением об ошибке, проверьте, удовлетворяют ли их значения заданному типу данных.
4. Если, например, дата заполнения не удовлетворяет заданному формату или не попадает в заданный интервал, проверьте, правильно ли распознаны данные в поле *Дата заполнения*. Возможно, после внесения исправлений сообщение об ошибке исчезнет из соответствующей области.

## 7.4. Экспорт

После проверки распознанных данных оператор осуществляет экспорт пакета, нажав кнопку **Экспортировать**. Экспорт производится согласно настройкам шаблона документов.

Если вы не хотите использовать настройки шаблона, вы можете произвести экспорт в файл или базу данных, выбрав соответствующий пункт выпадающего меню кнопки **Экспортировать**. При этом вы можете задавать любые настройки экспорта.

**Произведите экспорт данных в файл согласно произведенным ранее настройкам. Для этого достаточно нажать кнопку *Экспортировать*. Откройте полученный файл и просмотрите результаты экспорта. Пришлите этот файл в компанию АБВУУ по адресу [FlexiCapture\\_Feedback@abbyy.com](mailto:FlexiCapture_Feedback@abbyy.com).**

**Благодарим вас за проделанную работу и желаем успехов!**

## 8. Заключение

На простом примере мы разобрали все этапы настройки программы и обработки структурированных документов, но возможности программы шире, с ее помощью можно обрабатывать как простые, так и сложные многостраничные документы разных типов: полуструктурированные, неструктурированные, а также документы смешанных типов. При возникновении вопросов, обращайтесь к контекстной справке программы и руководству по установке.