

Система оптического распознавания текстов

ABBYY FineReader®

Версия 7.0

Руководство пользователя

Информация, содержащаяся в этом документе, может быть изменена без предварительного уведомления, и компания АBBYY не берет на себя на этот счет никаких обязательств.

Программное обеспечение, описываемое в этом документе, поставляется в соответствии с Лицензионным договором. Это программное обеспечение может быть использовано или скопировано лишь в строгом соответствии с условиями этого договора. Копирование этого программного обеспечения на какой-либо носитель информации, если на это нет специального разрешения в Лицензионном договоре или в соглашении о нераспространении, является нарушением Закона Российской Федерации “О правовой охране программ для ЭВМ и баз данных” и норм международного права.

Никакая часть настоящего Руководства ни в каких целях не может быть воспроизведена или передана в какой бы то ни было форме и какими бы то ни было средствами, будь то электронные или механические, включая фотокопирование и запись на магнитные носители, если на то нет письменного разрешения компании АBBYY.

© 2003, АBBYY Software Ltd. Все права защищены.

АBBYY, FineReader, Lingvo, Scan&Read, Scan&Translate, “принцип одной кнопки”, “Ваш компьютер читает сам”, “Ваш компьютер читает и переводит сам” – зарегистрированные торговые марки АBBYY; Try&Buy, DOCFLOW – торговые марки АBBYY.

Adobe®, Adobe Logo, Adobe PDF (Portable Document Format) и Adobe Acrobat® являются зарегистрированными торговыми марками компании Adobe Systems Incorporated.

Остальные торговые марки являются торговыми марками или зарегистрированными торговыми марками своих законных владельцев.

127015, Москва, а/я 72. АBBYY.

Содержание

| | |
|--|-----------|
| Добро пожаловать! | 7 |
| Глава 1 | |
| Установка и запуск | 9 |
| Требования к системе | 10 |
| Установка программы FineReader | 10 |
| Установка ABBYY FineReader Corporate Edition в сети | 11 |
| Об активации ABBYY FineReader | 12 |
| Глава 2 | |
| Быстрое знакомство | 15 |
| Как ввести документ за минуту | 16 |
| Главное окно программы FineReader | 16 |
| Инструментальные панели программы FineReader | 18 |
| Глава 3 | |
| Возможности программы ABBYY FineReader | 25 |
| Что такое OCR-системы | 26 |
| Новые возможности ABBYY FineReader 7.0 | 27 |
| Форматы, в которых FineReader сохраняет распознанный текст | 30 |
| оддерживаемые графические форматы | 31 |
| Глава 4 | |
| Получение и обработка изображения программой FineReader | 33 |
| Сканирование | 34 |
| Установка параметров сканирования | 35 |
| Советы по подбору яркости | 36 |
| Сканирование многостраничных документов | 37 |
| Открытие файлов с изображениями | 39 |
| Передача изображений через папку на локальном диске, в локальной сети или на FTP сервере | 40 |
| Добавление в пакет изображений со сдвоенными страницами | 40 |
| Добавление в пакет изображений с визитными карточками | 41 |
| Нумерация страниц при добавлении в пакет | 42 |
| Проверка и корректирование полученного изображения | 42 |
| Опции хранения изображений в пакете | 45 |

Глава 5

Анализ макета страницы 47

| | |
|---|----|
| Общая информация по анализу макета страницы | 48 |
| Типы блоков | 48 |
| Опции автоматического анализа макета страницы | 49 |
| Выделение и редактирование блоков вручную | 52 |
| Ручной анализ таблицы | 55 |
| Использование шаблонов блоков | 56 |

Глава 6

Распознавание 57

| | |
|--|----|
| Общая информация по распознаванию | 58 |
| Язык распознавания | 59 |
| Тип печати входного текста | 60 |
| Другие опции распознавания | 61 |
| Распознавание в фоновом режиме | 62 |
| Распознавание с обучением | 62 |
| Обучение эталона | 64 |
| Редактирование эталона | 66 |
| Пользовательские языки и группы языков | 67 |
| Создание нового языка | 68 |
| Создание новой группы языков | 69 |

Глава 7

Проверка и редактирование текста 71

| | |
|---|----|
| Проверка распознанного текста | 72 |
| Опции проверки и редактирования текста | 73 |
| Пополнение и удаление слов из пользовательского словаря | 75 |
| Редактирование текста в редакторе FineReader | 76 |
| Редактирование таблиц | 80 |

Глава 8

Сохранение во внешние редакторы и форматы 81

| | |
|--|----|
| Общая информация по сохранению распознанного текста | 82 |
| Опции сохранения распознанного текста | 82 |
| Сохранение распознанного текста в форматах RTE, DOC и Word XML | 85 |
| Сохранение распознанного текста в формате PDF | 86 |
| Сохранение распознанного текста в формате HTML | 87 |
| Сохранение распознанного текста в формате PowerPoint | 88 |
| Сохранение изображения страницы | 89 |

Глава 9**Коллективная работа в сети 91**

Коллективная работа с одним пакетом на нескольких компьютерах 93

Коллективная работа с пользовательскими языками
и пользовательскими словарями 94Коллективная работа со вспомогательными словарями к языкам
со словарной поддержкой 95**Глава 10****Работа с пакетами 97**

Общая информация по работе с пакетом 98

Создание нового пакета 99

Открытие пакета 99

Добавление изображений в пакет 100

Нумерация страниц пакета 100

Сохранение пакета 102

Закрытие страницы пакета или всего пакета 102

Удаление пакета 102

Настройки пакета 102

Поиск текста по распознанным страницам пакета 103

Приложение**Горячие клавиши и глоссарий 105**

Горячие клавиши 106

Глоссарий 109

Добро пожаловать!

Спасибо, что Вы приобрели программу ABBYY FineReader!

Несмотря на то, что в нашу жизнь все больше проникают электронные документы, бумажные источники информации по-прежнему не теряют своей актуальности. Все мы читаем книги, газеты и журналы, получаем факсы, письма, отчеты, материалы с выставок и множество других печатных документов. Как было бы здорово, если бы вся нужная информация немедленно попадала в компьютер и была под рукой, когда это необходимо! Ведь в современном стремительном мире более успешным будет тот, кто умеет управлять информацией, независимо от того, в каком виде она представлена. Как было бы здорово, если бы компьютер умел читать! Теперь у Вас есть средство, которое сделает эту мечту реальностью.

ABBYY FineReader – это система оптического распознавания текстов (Optical Character Recognition, OCR). Она предназначена для того, чтобы быстро перевести отсканированные бумажные документы, а также PDF-файлы в удобный для дальнейшего использования электронный формат – файл Word или Excel, HTML-страничку, презентацию PowerPoint или в любой другой из поддерживаемых форматов, при этом полностью сохранив оформление документа. Использование таких систем уже стало стандартом при работе с документами по всему миру, и все меньшему числу людей приходится перепечатывать бумажный документ, чтобы получить его электронный аналог.

Систему ABBYY FineReader очень легко использовать – даже тот, кто впервые открыл программу, сможет получить результат за несколько минут. При этом FineReader удовлетворяет самым высоким требованиям профессиональных пользователей благодаря большому количеству различных опций и настроек.

Это руководство поможет вам быстро и эффективно добиться отличного результата, используя все возможности продукта, и получить настоящее удовольствие от своей работы.

К делу!

Руководство пользователя

Данное руководство описывает основные действия системы распознавания. Каждая глава начинается с краткого описания и содержания.

Справочный файл

Справочный файл описывает основные действия системы распознавания и содержит подробную информацию о возможностях программы, различных установках, командах меню и диалогов.

Файл Readme

Файл Readme содержит общую информацию, появившуюся после выхода в свет печатных материалов к данному программному продукту.

Техническая поддержка

Если при использовании системы у Вас возникли вопросы, то прежде чем обратиться в отдел технической поддержки пользователей, просмотрите всю имеющуюся у Вас документацию (Руководство пользователя и справочный файл), а также зайдите на наш сайт в раздел технической поддержки www.abbyy.ru/support/index.htm – возможно, Вы найдете ответ на свой вопрос. Для того чтобы дать Вам квалифицированные рекомендации, работникам отдела поддержки пользователей необходимо иметь следующую информацию:

- Серийный номер дистрибутива (кроме пользователей Trial версий).
- Модель Вашего сканера.
- Общее описание проблемы с полным текстом сообщения об ошибке (если такое имеется).
- Версию системы Windows.
- Другую информацию, которую Вы считаете важной.

Часть этой информации можно получить автоматически, нажав на кнопку **System Info** в диалоге **О системе ABBYY FineReader** (меню **Справка/О программе**).

Глава 1

Установка и запуск

В данном разделе описаны основные инструкции по установке программы FineReader: системные требования, установка на одно рабочее место и сетевая установка.

Для установки FineReader используется специальная программа установки, входящая в состав дистрибутива. Всегда производите установку, используя CD–Rom только из дистрибутива, который Вы получили при покупке.

Краткое содержание раздела:

- Требования к системе
- Установка программы FineReader
- Установка ABBYY FineReader Corporate Edition в сети
- Запуск программы FineReader
- Об активации ABBYY FineReader

Требования к системе

1. ПК с процессором семейств Intel® Pentium®/Celeron®/Itanium®/Xeon™, AMD K6/Athlon™/Duron™ или совместимым с ними процессором, тактовая частота которого составляет не менее 200 МГц, или более мощным.
2. Операционная система Microsoft Windows XP, Windows 2000, Windows NT 4.0 (SP6 или выше), Windows Me/98 (для работы с локализованным интерфейсом операционная система должна обеспечивать необходимую языковую поддержку).
3. Требования к оперативной памяти зависят от используемой операционной системы: для Windows XP/2000/NT 4.0 – 64 Мб; Windows Me/98 – 32 Мб. При работе в многопроцессорных системах требуется дополнительно 16 Мб оперативной памяти для каждого дополнительного процессора.
4. Свободное место на диске: 150 Мб для обычной установки, 70 Мб для работы системы.
5. 100% Twain-совместимый сканер, цифровая камера или факс-модем.
6. Видеоплата и монитор с разрешением не менее 800×600 точек.
7. Клавиатура, мышь или другое указательное устройство.

Замечание: Для поиска текста по распознанным страницам пакета, просмотра канала новостей ABBYY Community (только для версии ABBYY FineReader 7.0 Professional Edition) необходим браузер Microsoft Internet Explorer 4.0 или его более поздняя версия.

Установка программы FineReader

Программа установки поможет Вам установить ABBYY FineReader. Перед началом установки рекомендуется закрыть все другие приложения.

Чтобы установить ABBYY FineReader:

1. Вставьте компакт-диск в дисковод для компакт-дисков. Программа установки запустится автоматически.
2. Следуйте инструкциям программы установки.

В случае если программа установки не запустилась автоматически:

1. Нажмите кнопку **Пуск** на **Панели Задач** и выберите пункт **Настройка/Панель управления**.
2. Дважды нажмите на значок **Установка и удаление программ**.
3. Выберите закладку **Установка и удаление** и нажмите кнопку **Установить...**
4. Далее следуйте инструкциям программы установки.

Опции установки

В процессе установки программы FineReader Вам будет предложено выбрать одну из предложенных опций установки:

- **Обычная** (рекомендуется) – устанавливаются все компоненты дистрибутива, в том числе все языки распознавания, язык интерфейса (один) – выбранный при установке.
- **Установка компонентов дистрибутива по выбору** – из предложенного набора компонентов, входящих в дистрибутив, Вы можете выбрать только те, которые нужны Вам (в том числе доступные языки распознавания).

Внимание! Если Вы хотите применять пользовательские словари и эталоны из предыдущей версии программы, не деинсталлируйте ее до инсталляции новой версии. Установив новую версию, Вы сможете **применять эталоны и словари предыдущей версии**.

Если при установке программы возникло сообщение об ошибке, обратитесь к файлу `readme.htm` (он находится на компакт-диске ABBYY FineReader 7.0), в котором приведены рекомендации, помогающие в большинстве случаев.

Установка ABBYY FineReader Corporate Edition в сети (только для администраторов сети)

Установка ABBYY FineReader 7.0 Corporate Edition на сетевой сервер осуществляется администратором сети. Установка программы ABBYY FineReader 7.0 Corporate Edition осуществляется в два этапа. Сначала программа устанавливается на сервер, а затем с сервера – на рабочие станции одним из четырех способов:

- используя Active Directory;
- с помощью Microsoft System Management Service (SMS);
- с помощью вызова командной строки;
- вручную, в интерактивном режиме.

Чтобы установить ABBYY FineReader 7.0 Corporate Edition на сервер:

1. Вставьте компакт-диск в дисковод для компакт дисков.
2. Используя командную строку, запустите файл `setup.exe` с параметром `/a` с установочного компакт-диска FineReader.

Подробную информацию по установке программы FineReader на рабочие станции, о работе с Менеджером лицензий, а также особенностях работы с программой в корпоративной сети Вы можете найти в “Руководстве системного администратора”, которое находится в подпапке **Administrator's Guide** папки сервера, куда был установлен FineReader.

Запуск программы FineReader

Чтобы запустить программу FineReader:

- Выберите пункт **ABBYY FineReader 7.0 Professional Edition (Corporate Edition)** в меню **Пуск/Программы**.

Внимание! Перед запуском программы убедитесь, что Ваш сканер подключен к компьютеру и включен в сеть. Если Ваш сканер еще не подключен, обратитесь к руководству пользователя сканера.

Если у Вас нет сканера, Вы можете использовать программу FineReader для распознавания графических файлов (примеры таких файлов находятся в папке *ABBYY FineReader/Demo*).

Об активации ABBYY FineReader

Компания ABBYY принимает меры по борьбе с компьютерным пиратством и защите прав на интеллектуальную собственность. Пиратство причиняет ущерб не только производителям программного обеспечения, но и конечным пользователям. Приобретая лицензионные продукты, пользователи получают программные продукты без каких-либо изменений, сделанных третьими лицами, в то время как пиратские продукты не могут гарантировать этого.

Начиная с версии ABBYY FineReader 7.0, в состав продукта включается специальная технология защиты программного обеспечения. Она призвана исключить возможность использования копий продуктов в случае отсутствия у пользователя лицензионного соглашения с правообладателем на право использования данного программного продукта.

В случае если лицензионное соглашение разрешает установить и использовать программный продукт только на одном персональном компьютере, установка этого же продукта на другие персональные компьютеры приводит к нарушению условий лицензионного соглашения, а также к нарушению законодательства РФ о защите авторских прав. Технология активации позволяет технически ограничить число используемых рабочих копий продукта и исключить установку одной лицензионной копии на неограниченное число персональных компьютеров. При этом число установок и активаций программы на одном и том же персональном компьютере не ограничивается.

В зависимости от версии продукта и территории его распространения, возможны следующие ограничения по функциональности:

- Невозможность сохранения и печати распознанного текста на языках с кириллическим алфавитом (**ABBYY FineReader 7.0 Professional Edition**);
- Невозможность сохранения и печати распознанного текста на всех поддерживаемых языках (**ABBYY FineReader 7.0 Professional Edition**);
- Версия **ABBYY FineReader 7.0 Corporate Edition** без активации не работает.

Как осуществляется активация?

Процесс активации занимает немного времени и осуществляется при помощи специального **Мастера активации**, который встроен в программу. Мастер активации имеет простой интерфейс, с помощью которого пользователь передает необходимую для активации информацию в компанию АБВУУ, а затем вводит в программу полученный Активационный код (для версии Professional Edition) или Активационный файл (для версии Corporate Edition).

Информация для активации передается в виде кода – **Installation ID**, который формируется на основе информации о компьютере пользователя, на который происходит установка программы. При создании кода не используются персональные сведения о пользователе или о его компьютере, которые позволили бы идентифицировать пользователя.

Возможные способы активации:

- **Активация через Интернет.**
Процесс активации осуществляется автоматически и занимает несколько секунд. Для этого способа активации требуется наличие соединения с Интернетом.
- **Активация по электронной почте.**
Пользователю предлагается отправить в компанию АБВУУ автоматически сформированное письмо с информацией, необходимой для активации программы. Пожалуйста, не редактируйте текст и тему письма, чтобы получить быстрый ответ от почтового робота.
- **Активация по телефону или факсу** (только для версии Professional Edition).
Пользователю предлагается позвонить в ближайший офис компании АБВУУ или партнеру АБВУУ и продиктовать Installation ID и серийный номер программы оператору. Для большинства стран возможна активация по факсу. Сообщение, с информацией, необходимой для активации программы будет сформировано автоматически. Пользователю будет предложено распечатать и отправить его по указанному факсу, предварительно вписав свой номер факса, по которому будет выслан Активационный код.

При активации через Интернет процесс осуществляется автоматически. При активации по телефону, факсу или электронной почте пользователь вводит полученный от ABBYY Активационный код или Активационный файл в соответствующие поля диалогового окна мастера активации.

После завершения процесса активации продукт готов к постоянному использованию на данном компьютере.

ABBY FineReader 7.0 можно переустанавливать на одном и том же компьютере без повторной активации столько раз, сколько потребуется пользователю. Однако, если перед переустановкой программы была существенно изменена конфигурация компьютера, отформатирован жесткий диск или переустановлена операционная система, потребуется повторная активация программы с получением нового активационного кода.

Защита конфиденциальности данных пользователей

Для использования ABBYY FineReader в полнофункциональном режиме требуется активация продукта. В процессе активации в ABBYY не передается никаких персональных данных пользователя и сохраняется его полная анонимность.

Мастер активации FineReader создает уникальный параметр Installation ID. Этот параметр создается на основе данных о конфигурации компьютера в момент активации. Этот параметр не включает в себя никакой персональной информации, а также никакой информации о модели компьютера, об установленных на нем программных продуктах, пользовательских настройках или хранящихся на компьютере данных. Помимо Installation ID, в процессе активации на сервер компании ABBYY передается серийный номер, название, номер версии, а также язык интерфейса FineReader. Эта информация используется только для того, чтобы правильно выбрать язык и содержание сообщения о результатах активации, и не будет использована ни в каких других целях.

Небольшие изменения в конфигурации компьютера не потребуют проведения повторной активации. Однако при значительных изменениях, например, при полной замене компьютера или переустановке операционной системы, потребуется повторная активация продукта.

Глава 2

Быстрое знакомство

В этой главе Вы узнаете, как можно ввести нужный документ в компьютер, практически ничего не зная о системе FineReader. Также в этой главе приводится обзор основных элементов системы FineReader, таких как окна системы, панели инструментов, строка состояния.

Краткое содержание раздела:

- Как ввести документ за минуту
- Главное окно программы FineReader
- Инструментальные панели программы FineReader

Как ввести документ за минуту

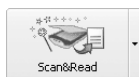
1. Включите сканер (если он имеет отдельный от компьютера источник питания).

Внимание! Многие модели сканера необходимо включать до включения компьютера.

2. Включите компьютер и запустите FineReader (**Пуск/Программы/ABBYY FineReader 7.0 Professional Edition** или **Corporate Edition**).

Перед Вами откроется Главное окно программы FineReader.

3. Вставьте в сканер страницу, которую Вы хотите распознать.
4. Нажмите на стрелку справа от кнопки **Scan&Read**, в открывшемся локальном меню выберите пункт **Мастер Scan&Read**.



Мастер Scan&Read вызывает специальный режим, при котором Вы можете отсканировать и распознать страницу или открыть и распознать графическое изображение (пример графического файла Вы можете найти в папке **Demo**. Она находится в папке, куда Вы установили FineReader). При этом каждый шаг сопровождается подсказками системы.

5. Далее следуйте указаниям **Мастера Scan&Read**.

Процесс ввода документов в компьютер состоит из четырех этапов: *сканирование, распознавание, проверка и сохранение* результатов распознавания.

В результате сканирования появится окно **Изображение**, содержащее “фотографию” страницы. Затем программа попросит Вас установить параметры распознавания и приступит к распознаванию изображения, одновременно анализируя его. Обработанные участки изображения закрашиваются голубым цветом.

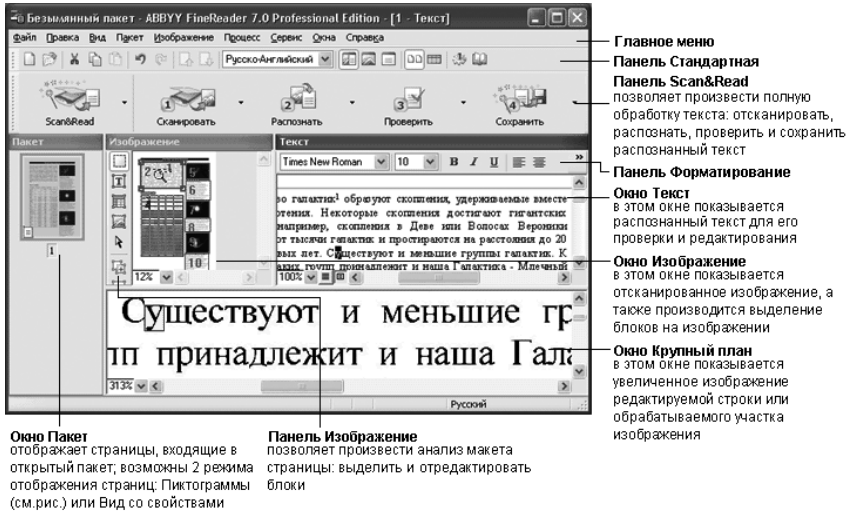
Результат распознавания Вы увидите в окне **Текст**. В этом же окне Вы можете проверить и отредактировать распознанный текст. Следуя далее указаниям **Мастера Scan&Read**, Вы можете либо передать распознанный текст в выбранное Вами приложение или сохранить его на диск, либо продолжить обработку следующих изображений.

Главное окно программы FineReader

Работая с программой FineReader, пользователь всегда имеет дело с некоторым пакетом. Пакет – это папка, в которой хранятся изображения и рабочие файлы программы. Каждое отсканированное изображение записывается как отдельная страница пакета.

При запуске программы FineReader по умолчанию открывается новый пакет. Вы можете продолжить работу с новым пакетом или же открыть уже существующий пакет. Вы мо-



жете продолжить работу с пакетом по умолчанию или создать новый пакет. Более подробно о работе с пакетом см. Общие сведения по работе с пакетом.



Вверху **Главного окна** FineReader находится меню системы, под ним – инструментальные панели. В программе их четыре: Стандартная, Форматирование, Изображение и Scan&Read.

Спрятать или показать инструментальные панели на экране можно через меню **Вид** (пункт **Панели инструментов**) или через локальное меню. Чтобы открыть локальное меню, щелкните правой кнопкой мыши на одной из инструментальных панелей. Панели, которые видны на экране, отмечены галочкой. Выберите панель, которую Вы хотите спрятать или, наоборот, показать.

Чтобы выбрать вид окна Пакет:

- Нажмите на соответствующую кнопку ( / ) на панели инструментов **Стандартная**, или
- Нажмите правой кнопкой мыши на окне **Пакет** и выберите в локальном меню пункт **Вид>...**

Внизу окна расположена информационная панель (еще ее называют строкой состояния), где Вы можете получить сведения о состоянии системы и производимых ею операциях, а также краткую справку о выбираемых Вами пунктах меню и кнопках.

Остальное пространство Главного окна занимают по мере своего появления рабочие окна системы: **Пакет**, **Изображение**, **Крупный план** и **Текст**.

Окна **Изображение**, **Крупный план** и **Текст** связаны между собой: при двойном щелчке на изображении в окне **Изображение** курсор в окнах **Крупный план** и **Текст** (при наличии распознанного текста) переместится на ту же позицию, что и в окне **Изображение**.

Вы можете поменять взаимное расположение окон на экране:

- В меню **Вид** выберите один из пунктов: **Окно Пакет>...; Окна Изображение и Текст>...; Окно Крупный план>...**

| Вариант расположения окон: | Применение: |
|---|--|
| Окно Пакет слева; тип отображения страниц пакета: Пиктограммы ; Окна Изображение , Текст и Крупный план | Используется при малом количестве страниц пакета. |
| Окно Пакет сверху; тип отображения страниц пакета: Вид со свойствами ; Окна Изображение , Текст и Крупный план . | Используется при большом количестве страниц пакета. |
| Окно Пакет сверху; тип отображения страниц пакета: Вид со свойствами ; Окна Изображение и Крупный план . | Используется для анализа и распознавания большого количества страниц пакета. |
| Окно Пакет сверху; тип отображения страниц пакета: Вид со свойствами ; Окна Текст и Крупный план . | Используется при редактировании большого количества страниц пакета. |

Клавиши для работы с окнами

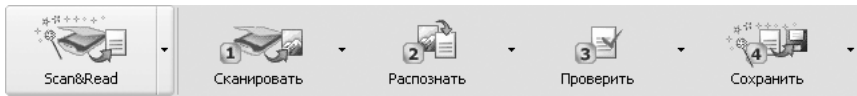
- Для переключения между окнами нажмите **Ctrl+TAB**.
- Чтобы активизировать окно **Пакет**, нажмите **Alt+1**.
- Чтобы активизировать окно **Изображение**, нажмите **Alt+2**.
- Чтобы активизировать окно **Текст**, нажмите **Alt+3**.

Инструментальные панели программы FineReader

В системе FineReader инструментальных панелей всего 4: **Стандартная**, **Изображение**, **Форматирование** и главная панель программы – **Scan&Read**. Кнопки на инструментальных панелях – самое удобное средство доступа к операциям системы. Те же операции можно выполнять из меню программы или с помощью горячих клавиш.

Если Вы хотите узнать о назначении той или иной кнопки на инструментальной панели, установите на ней курсор мыши. Под кнопкой появится подпись (tooltip), а на информационной панели будет выведено более подробное сообщение о назначении этой кнопки.

Главная панель программы – Scan&Read

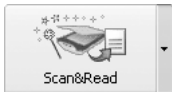


Кнопки на панели **Scan&Read** связаны с базовыми операциями системы – **Сканированием, Распознаванием, Проверкой и Сохранением** результатов распознавания. Цифры на кнопках указывают, в каком порядке нужно выполнить действия, чтобы получить электронную версию бумажного документа. Каждое из этих действий можно провести по отдельности или объединить в одно, нажав на кнопку **Мастер Scan&Read**. Она позволяет провести полный цикл обработки текста автоматически. Каждая из кнопок имеет несколько режимов работы. Нажав на стрелку справа от кнопки, в открывшемся локальном меню Вы можете выбрать один из них, при этом “информация” об этом отразится на значке кнопки. Для того чтобы повторить ту же операцию для другого изображения, Вам достаточно повторно нажать кнопку.

Название кнопки

Режимы работы

Scan&Read



Сканировать и распознать
Сканировать и распознать несколько страниц
Открыть и распознать
Мастер Scan&Read...

Мастер Scan&Read – запускает специальный режим сканирования и распознавания, во время которого система контролирует действия пользователя и подсказывает ему, что надо делать, чтобы получить тот или иной результат.

Сканировать и распознать – запускает сканирование и распознавание документа.

Сканировать и распознать несколько страниц – сканирует и распознает несколько страниц в цикле.

Открыть и распознать – позволяет открыть и распознать изображения, выбранные в диалог **Открыть (Open)**.

1–Сканировать



Открыть изображение...
Сканировать изображение...
Сканировать несколько страниц...
Просмотр папки...
Опции...

Открыть изображение – добавляет изображение в пакет, при этом копия изображения сохраняется в папке пакета.

Сканировать изображение – сканирует изображение.

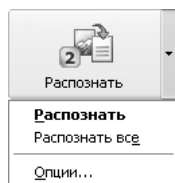
Сканировать несколько страниц – сканирует изображения в цикле. Чтобы остановить сканирование, в меню **Файл** выберите пункт **Остановить сканирование**.

Просмотр папки (только в версии Corporate Edition) – позволяет запустить специальный режим мониторинга заданной папки, при котором добавленные в папку

изображения автоматически открываются в окне программы. Чтобы отключить этот режим, в меню **Файл** выберите пункт **Отключить режим просмотра папки**.

Опции – открывает закладку Сканирование/Открытие диалога **Опции**, на которой Вы может установить опции сканирования и предварительной обработки документа.

2–Распознать



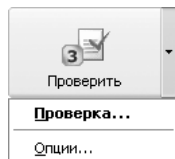
Распознать – распознает открытую страницу (или выделенные страницы) пакета.

Распознать все – распознает все нераспознанные страницы пакета.

Опции – открывает закладку Распознавание диалога

Опции, на которой Вы может установить опции распознавания документа.

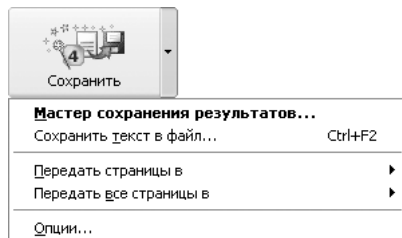
3–Проверить



Проверить – позволяет найти в тексте слова, содержащие неуверенно распознанные символы, и неправильно написанные слова.

Опции – открывает закладку Проверка диалога **Опции**, на которой Вы можете установить опции проверки документа.

4–Сохранить



Мастер сохранения результатов – открывает диалог **Мастер сохранения результатов**, в котором Вы можете выбрать приложение для сохранения и установить опции сохранения.

Сохранить текст в файл – сохраняет распознанный текст в файл на диск.

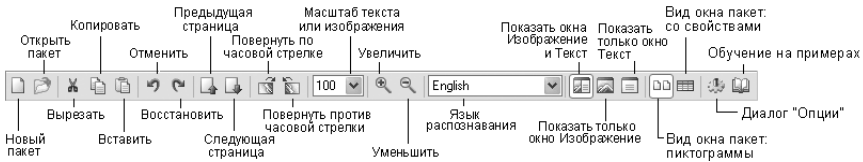
Передать страницы в – напрямую передает распознанный текст в выбранное приложение без сохранения его на диск. При передаче распознанного текста с нескольких страниц пакета сначала выделите их в окне Пакет.

Передать все страницы в – передает все распознанные страницы в выбранное приложение без сохранения их на диск.

Опции – открывает закладку Форматирование диалога **Опции**, на которой Вы можете установить опции сохранения документа.

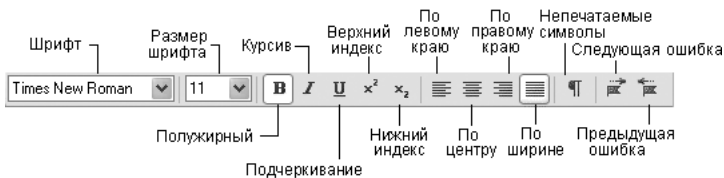
Панель Стандартная

На панели **Стандартная** находятся кнопки, управляющие работой с файлами и изображением (отмена и повтор действия, перемещение по страницам пакета, очистка и поворот изображения), а также список языков распознавания.



Замечание: Вид окна программы FineReader, точнее количество кнопок на панелях инструментов **Изображение**, **Стандартная** и **Форматирование**, зависит от разрешения экрана Вашего компьютера. Чтобы увидеть все кнопки, упомянутые в руководстве и справке, поменяйте разрешение экрана на более высокое. Поскольку кнопки реализуют лишь часть доступных в программе FineReader команд и дублируются опциями меню, отсутствие нескольких кнопок на панелях не ограничивает функциональность программы.

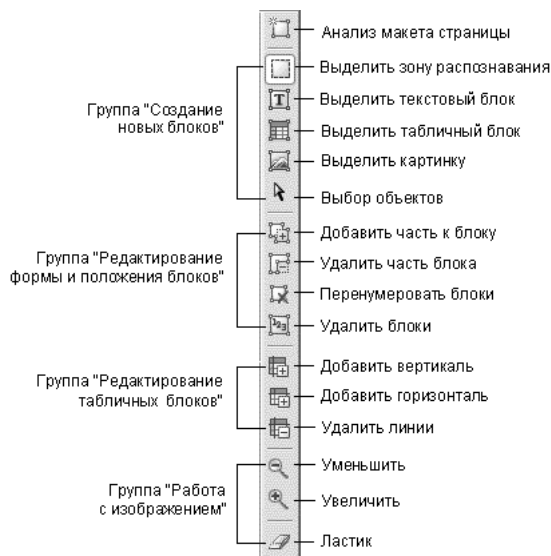
Панель Форматирование



На панели **Форматирование** находятся кнопки, позволяющие изменить оформление текста.

Панель Изображение

Панель **Изображение** содержит кнопки, позволяющие производить анализ макета страницы (например, создать и отредактировать блоки), а также кнопки, позволяющие увеличить/уменьшить масштаб изображения, отредактировать изображение (например, стереть ненужные участки изображения, такие, как подписи или большие участки мусора).



Замечание: Кнопки, позволяющие создать и отредактировать блоки, можно использовать не только в окне **Изображение**, но и в окне **Крупный план**.

Настройка инструментальных панелей

FineReader позволяет настраивать инструментальные панели **Стандартная**, **Изображение** и **Форматирование**: добавлять и удалять кнопки, осуществляющие доступ ко всем командам программы.

Каждому пункту меню соответствует свой значок. Полный список команд и соответствующих им кнопок приведен в диалоге **Настройка** (меню **Сервис>Настройка**), в списке **Команды**.

Чтобы добавить кнопку на панель инструментов:

1. В поле **Категории** выберите нужную категорию.

Замечание: Список категорий команд сформирован по названию меню. Выбор категории меняет набор команд в списке **Команды**.

2. В поле **Панели инструментов** выберите инструментальную панель, на которую требуется добавить команду.
3. В списке **Команды** выберите команду и нажмите кнопку (>>).

Выбранная команда появится в списке команд выбранной инструментальной панели и на панели в главном окне программы.

Чтобы удалить кнопку с панели инструментов:

- В списке **Кнопки панели инструментов** выберите кнопку, которую Вы хотите удалить, и нажмите кнопку (<<).

Замечания:

1. Порядок расположения кнопок в списке определяет их порядок на панели инструментов. Чтобы изменить порядок кнопок, выделите команду в списке команд текущей инструментальной панели и нажмите кнопки **Вверх (Вниз)** для перемещения команды вверх (вниз) по списку.
 2. Вы можете разделить команды по группам. Для этого в списке **Команды** выберите пункт **Разделитель** и нажмите кнопку **Добавить**. Разделитель добавится в список кнопок текущей панели инструментов. Если требуется, переместите его.
 3. Чтобы восстановить стандартные кнопки на панели инструментов, выбранной в списке **Панели инструментов**, нажмите кнопку **Восстановить**. Чтобы восстановить стандартные кнопки на всех инструментальных панелях, нажмите кнопку **Восстановить все**.
-

Глава 3

Возможности программы ABBYY FineReader

FineReader позволяет ввести документ одним нажатием на кнопку Scan&Read, не вдаваясь в подробности работы программы. Распознанный текст можно передать в текстовый редактор или электронную таблицу, сохранить в форматах RTE, DOC, Word XML (при сохранении в Microsoft Office 2003), PDF и HTML с полным сохранением оформления документа или сохранить в базе данных.

Краткое содержание раздела:

- Что такое OCR–системы
- Новые возможности ABBYY FineReader 7.0
- Форматы, в которых FineReader сохраняет распознанный текст
- Поддерживаемые графические форматы

Что такое OCR-системы

Системы оптического распознавания символов (Optical Character Recognition – OCR) предназначены для автоматического ввода печатных документов в компьютер.

FineReader — omnifонтовая система оптического распознавания текстов. Это означает, что она позволяет распознавать тексты, набранные практически любыми шрифтами, без предварительного обучения. Особенностью программы FineReader является высокая точность распознавания и малая чувствительность к дефектам печати, что достигается благодаря применению технологии “целостного целенаправленного адаптивного распознавания”.

Процесс ввода документа в компьютер можно подразделить на два этапа:

1. **Сканирование.** На первом этапе сканер играет роль “глаз” Вашего компьютера: “просматривает” изображение и передает его компьютеру. При этом полученное изображение является не чем иным, как набором черных, белых или цветных точек, картинкой, которую невозможно отредактировать ни в одном текстовом редакторе.
2. **Распознавание.** Обработка изображения OCR-системой.

Остановимся на втором шаге более подробно.

Обработка изображения системой FineReader включает в себя анализ графического изображения, переданного сканером, и распознавание каждого символа. Процессы анализа макета страницы (определение областей распознавания, таблиц, картинок, выделение в тексте строк и отдельных символов) и распознавания изображения тесно связаны между собой: алгоритм поиска блоков использует информацию о распознанном тексте для более точного анализа страницы.

Как уже упоминалось, распознавание изображения осуществляется на основе технологии “целостного целенаправленного адаптивного распознавания”.

- **Целостность** – объект описывается как целое с помощью значимых элементов и отношений между ними.
- **Целенаправленность** – распознавание строится как процесс выдвижения и целенаправленной проверки гипотез.
- **Адаптивность** – способность OCR-системы к самообучению.

В соответствии с этими тремя принципами система сначала выдвигает гипотезу об объекте распознавания (символе, части символа или нескольких склеенных символах), а затем подтверждает или опровергает ее, пытаясь последовательно обнаружить все структурные элементы и связывающие их отношения. В каждом структурном элементе выделяются части, значимые для человеческого восприятия: отрезки, дуги, кольца и точки.

Следуя принципу адаптивности, программа самостоятельно “настраивается”, используя положительный опыт, полученный на первых уверенно распознанных символах. Целе-направленный поиск и учет контекста позволяют распознавать разорванные и искаженные изображения, делая систему устойчивой к возможным дефектам письма.

В результате работы в окне FineReader появится распознанный текст, который Вы можете отредактировать и сохранить в наиболее удобном для Вас формате.

Новые возможности ABBYY FineReader 7.0

Точность распознавания

- Точность распознавания улучшена на 25%. Лучше анализируются и распознаются документы сложной верстки, в частности, содержащие участки текста на цветном фоне или фоне, состоящем из мелких точек, документы со сложными таблицами, в том числе таблицами с белыми разделителями, таблицами с цветными ячейками
- В новую версию добавлены специализированные словари для английского и немецкого языка, включающие наиболее часто используемые юридические и медицинские термины. Это позволяет достичь качественно нового уровня при распознавании документов юридической и медицинской тематики.

Поддержка формата XML и интеграция с Microsoft Office

- В FineReader появился новый формат сохранения – Microsoft Word XML. Теперь пользователи новой версии Microsoft Office 2003 смогут работать с документами, распознанными FineReader, используя все преимущества формата XML!
- Интеграция FineReader с Microsoft Word 2003 позволяет объединить мощные возможности этих двух приложений для обработки распознанного текста. Вы сможете проверять и редактировать результаты распознавания с помощью привычных инструментов Word, одновременно сверяя переданный в Word текст с оригинальным изображением – окно Zoom FineReader открывается прямо в окне Word.
- Новые возможности сделают Вашу работу более удобной. При создании документа Word можно вызвать FineReader, распознать текст и вставить его в то место документа, где находится курсор, то есть Вы сможете легко собрать в одном документе информацию из разных бумажных источников или PDF-файлов. Результаты распознавания теперь можно отправить по электронной почте как вложение в любом из поддерживаемых форматов сохранения.

Улучшена работа FineReader с PDF документами

- Качество распознавания PDF-файлов значительно улучшилось. Большинство документов, помимо изображения страницы, содержат текст. FineReader 7.0 умеет извлекать этот текст и использовать его для проверки результатов и улучшения качества распознавания.
- Теперь Вы можете редактировать распознанные PDF-документы в окне редактора FineReader: внесенные изменения будут сохранены в любом из поддерживаемых в программе режимов сохранения PDF-файлов.
- Формат PDF-файлов, создаваемых программой FineReader, оптимизирован для их публикации в Интернете – пользователь сможет просматривать содержание первых страниц, пока скачивается остальная часть документа.

Новые возможности сохранения

- Новый формат сохранения результатов распознавания – **Microsoft PowerPoint** – позволяет быстро создавать новые презентации или редактировать уже существующие.
- При сохранении в **Microsoft Word** уменьшился размер получаемого файла, улучшилось сохранение форматирования документов с различными разделителями, появились новые опции сохранения картинок.
- Улучшено отображение сложных элементов верстки при сохранении в **HTML**, например, обтекание прямоугольных картинок текстом. Кроме того, уменьшился размера HTML-файла, что очень важно для публикации документов в Интернете.

Удобство использования

- Обновленный **интуитивно-понятный пользовательский интерфейс**. Стало удобнее работать с профессиональными настройками. Панели инструментов редактирования перенесены в окно, в котором отображаются результаты распознавания. Появились удобные инструменты управления окнами FineReader: например, можно поставить удобную степень увеличения в каждом из окон.
- Обновленное **практическое руководство** по улучшению качества распознавания поможет начинающему пользователю быстро начать работу, а более опытному – наилучшим образом настроить программу для получения отличного результата при работе с любыми типами документов.

Профессиональные возможности

Теперь в версии **FineReader Professional Edition** стали доступны те возможности, которые раньше были доступны только пользователям версии **Corporate Edition**:

- Улучшено распознавание **штрих-кодов**, поддержано распознавание двухмерного штрих-кода PDF-417.
- **Инструмент для разбиения изображения**. С помощью него Вы сможете разделить изображения на области и сохранить каждую область как отдельную страницу пакета. Таким образом удобно распознавать несколько визитных карточек, отсканированных вместе, книги, или распечатки слайдов презентаций PowerPoint.
- **Морфологический поиск**. Любой пакет, созданный в программе FineReader, можно использовать в качестве небольшой базы данных с возможностью полнотекстового морфологического поиска. Среди всех распознанных страниц пакета можно найти те страницы, которые содержат заданные слова во всех их грамматических формах (для 34 языков со словарной поддержкой).
- **Поддержка процессоров Intel, использующих технологию Hyper-Threading**. Использование этой технологии позволяет заметно увеличить производительность, что особенно важно, если стоит задача распознавания большого числа документов.

В версии FineReader 7.0 также появились и другие профессиональные возможности:

- **Двухстороннее сканирование**. Сканируя документ с напечатанным с двух сторон текстом с помощью поддерживающего эту опцию сканера, Вы получите изображения содержимого каждой стороны в виде двух отдельных страниц пакета. Если нужно отсканировать только одну сторону документа, эту опцию можно отключить.
- Поддержано открытие графических файлов формата **JPEG 2000** и сохранение в этот формат.

Сетевые возможности версии FineReader Corporate Edition

Подробно все особенности установки и использования FineReader Corporate Edition в корпоративной сети описаны в Руководстве системного администратора, который Вы можете найти в подпапке **Administrator's Guide** папки сервера, куда был установлен FineReader.

Основные улучшения по сравнению с предыдущей версией:

- **Поддержка основных способов автоматической установки с сервера на рабочие станции.** FineReader Corporate Edition поддерживает все основные способы автоматической установки в локальной сети: с помощью Active Directory, Microsoft Systems Management Server или с помощью командной строки.
- **Работа с многофункциональными устройствами, в том числе сетевыми.** Многофункциональные устройства, совмещающие в себе функции сканера, принтера, копира и факса, становятся все более популярными. Теперь необязательно ставить каждому сотруднику свой сканер – достаточно одного мощного устройства, с которым работают все пользователи организации. FineReader умеет работать с такими устройствами, как подключенными к рабочей станции, так и сетевыми. Специальные настройки программы позволяют пользователю автоматически открывать и отсканированные изображения с любого места в локальной сети или с ftp сервера и распознавать их
- **Различные модели корпоративного лицензирования.** Кроме лицензирования по числу одновременного работающих пользователей, стали доступны также другие способы лицензирования. Вы сможете выбрать тот вариант, который в большей степени отвечает Вашим потребностям.
- **License Manager** – средство управления лицензиями в сети. В **FineReader Corporate Edition** появилась удобная утилита управления лицензиями (Менеджер лицензий). С помощью нее осуществляется отслеживание использования FineReader на рабочих станциях, резервирование лицензий за рабочими станциями, добавление новых лицензий.

Форматы, в которых FineReader сохраняет распознанный текст

FineReader сохраняет результаты распознавания в следующих форматах:

- Microsoft Word Document (*.DOC)
- Rich Text Format (*.RTF)
- Microsoft Word XML Document (*.XML) (только для Microsoft Word 2003)
- Adobe Acrobat Format (*.PDF)
- HTML
- Microsoft PowerPoint Format (*.PPT)
- Comma Separated Values File (*.CSV)
- Простой текст (*.TXT). FineReader поддерживает различные кодовые страницы (Windows, DOS, Mac, ISO) и кодировку Unicode.
- Microsoft Excel Spreadsheet (*.XLS)
- DBF

оддерживаемые графические форматы

FineReader открывает файлы следующих форматов:

PDF:

Файлы в формате PDF, версии 1.3 или ранней

BMP:

2-битный – ч/б

4- и 8-битный – Palette

16-битный

24-битный – Palette и TrueColor

32-битный

PCX, DCX:

2-битный – ч/б

4- и 8-битный – Palette

24-bit – TrueColor

JPEG:

серый и цветной

JPEG 2000:

серый и цветной

TIFF:

ч/б – несжатый, CCITT3, CCITT3FAX, CCITT4, Packbits

серые – несжатый, Packbits, JPEG

TrueColor – несжатый, JPEG

Palette – несжатый, Packbits

многостраничный TIFF

PNG:

ч/б, серый, цветной

FineReader сохраняет изображения в следующих форматах:

BMP:

ч/б, серый, цветной

PCX:

ч/б, серый

JPEG:

серый, цветной

JPEG 2000:

серый и цветной

TIFF:

ч/б – несжатые, CCITT3, CCITT4, packbits

серые – несжатые, packbits, JPEG

цветные – несжатый и JPEG

PNG:

ч/б, серый, цветной

Глава 4

Получение и обработка изображения программой FineReader

Качество распознавания во многом зависит от качества исходного изображения. В этой главе Вы узнаете о том, как правильно отсканировать документ, как открыть и распознать уже имеющиеся на компьютере изображения (список поддерживаемых системой форматов приведен в разделе “Поддерживаемые графические форматы”), как можно обработать изображение и устранить некоторые его дефекты (например, мусор, возникший при сканировании) и т.д.

Краткое содержание раздела:

- Сканирование
- Установка параметров сканирования
- Советы по подбору яркости
- Сканирование многостраничных документов
- Открытие файлов с изображениями
- Передача изображений через папку на локальном диске, в локальной сети или на FTP сервере
- Добавление в пакет изображений со сдвоенными страницами
- Добавление в пакет изображений с визитными карточками
- Нумерация страниц при добавлении в пакет
- Проверка и корректирование полученного изображения
- Опции хранения изображений в пакете

Сканирование

FineReader работает со сканерами через TWAIN-интерфейс. Это единый международный стандарт, введенный в 1992 году для унификации взаимодействия устройств для ввода изображений в компьютер (например, сканера) с внешними приложениями. При этом возможно два варианта взаимодействия программы со сканерами через TWAIN-драйвер:

- **через интерфейс FineReader:** в этом случае для настройки опций сканирования используется диалог программы FineReader Настройки сканера;
- **через интерфейс TWAIN-драйвера сканера:** для настройки опций сканирования используется диалог TWAIN-драйвера сканера.

Преимущества одного режима перед другим

В режиме **Использовать интерфейс TWAIN-драйвера сканера**, как правило, доступна функция предварительного просмотра изображения (preview), позволяющая точно задать размеры сканируемой области, подобрать яркость, тут же контролируя результаты этих изменений. К сожалению, диалог TWAIN-драйвера сканера у каждого сканера выглядит по-своему, в большинстве случаев все надписи на английском языке. Вид этого окна и смысл опций описан в документации, прилагаемой к сканеру.

В режиме **Использовать интерфейс FineReader** доступны такие опции, как возможность сканирования в цикле на сканерах без автоподатчика, сохранение опций сканирования в отдельный файл **Шаблон пакета** (*.fbt) и возможность использования этих опций в других пакетах.

Вы можете легко переключаться между этими режимами:

- на закладке Сканирование/Открытие диалога **Опции** (меню **Сервис>Опции**) установите переключатель в одно из положений: **Использовать интерфейс TWAIN-драйвера сканера** или **Использовать интерфейс FineReader**.

Замечания:

1. Для некоторых моделей сканеров опция **Использовать интерфейс FineReader** может быть по умолчанию отключена (недоступна).
2. Чтобы в режиме **Использовать интерфейс FineReader** показывался диалог Настройки сканера, на закладке Сканирование/Открытие (**Сервис>Опции**) отметьте пункт **Запрашивать опции перед началом сканирования**.

Важно! Для того чтобы правильно подключить сканер, обратитесь к документации, прилагаемой к сканеру. При установке не забудьте установить **все** программное обеспечение, поставляемое вместе со сканером (драйвер TWAIN и/или сканирующую программу).

Чтобы запустить сканирование:



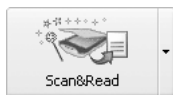
Нажмите кнопку **1–Сканировать** или в меню **Файл** выберите пункт **Сканировать**. Спустя некоторое время в **Главном окне** программы FineReader появится окно **Изображение** с “фотографией” вставленного листа.

Если Вы хотите отсканировать несколько страниц, то нажмите на стрелке справа от кнопки **1–Сканировать** и в локальном меню выберите пункт **Сканировать несколько страниц**.

В случае если сканирование не началось сразу:

- откроется встроенный **TWAIN–интерфейс** сканера;
- откроется диалог **Настройки сканера**.

Совет: Если Вы хотите сразу запустить распознавание отсканированных страниц, воспользуйтесь опцией **Сканировать и распознать** или **Сканировать и распознать несколько страниц**:



Нажмите на стрелке справа от кнопки **Scan&Read** и в локальном меню выберите один из пунктов: **Сканировать и распознать** или **Сканировать и распознать несколько страниц**.

FineReader отсканирует и распознает изображения. В **Главном окне** программы появятся окно **Изображение** с “фотографией” вставленного листа и окно **Текст** с результатом распознавания. Распознанный текст Вы можете сохранить во внешние редакторы и форматы.

Установка параметров сканирования

Качество распознавания во многом зависит от того, насколько хорошее изображение получено при сканировании. Качество изображения регулируется установкой основных параметров сканирования: типа изображения, разрешения и яркости.

Основными параметрами сканирования являются:

- **Тип изображения** – серый (256 градаций).
Сканирование в сером является оптимальным режимом для системы распознавания. В случае сканирования в сером режиме осуществляется автоматический подбор яркости. **Черно–белый** тип изображения обеспечивает более высокую скорость сканирования, но при этом теряется часть информации о буквах, что может привести к ухудшению качества распознавания на документах среднего и низкого качества печати. Если Вы хотите, чтобы содержащиеся в документе цветные элементы (картинки, цветные буквы и цветной фон) были переданы в электронный документ с сохранением цвета, необходимо выбрать **цветной** тип изображения. В других случаях используйте **серый** тип изображения.

- **Разрешение** – используйте 300 dpi для обычных текстов (размер шрифта 10 и более пунктов) и 400–600 dpi для текстов, набранных мелким шрифтом (9 и менее пунктов).
- **Яркость** – в большинстве случаев подходит среднее значение яркости – 50%.
На некоторых документах при сканировании в черно–белом режиме может понадобиться дополнительная настройка яркости.

Замечание: Сканирование с разрешением 400–600 dpi вместо 300 dpi или сканирование в сером или в цвете, а не в черно–белом может занять существенно больше времени. На некоторых моделях сканеров сканирование с разрешением 600 dpi занимает в 4 раза больше времени, чем сканирование с разрешением 300 dpi.

Чтобы установить параметры сканирования:

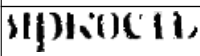
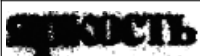
- При сканировании через TWAIN с использованием **интерфейса FineReader**: в меню **Сервис** выберите пункт **Настройки сканера** и в открывшемся диалоге Настройки сканера установите нужные опции.
- При сканировании через TWAIN с использованием **интерфейса TWAIN–драйвера сканера** для установки параметров сканирования используется диалог Вашего сканера, который открывается автоматически при нажатии на кнопку **1–Сканировать**. Опции для установки параметров сканирования могут называться по–разному, в зависимости от модели сканера. Например, яркость может называться brightness, threshold, изображаться “солнышком” или черно–белым кружком. Смысл опций описан в документации, прилагаемой к Вашему сканеру.

Советы по подбору яркости

Отсканированное изображение должно быть удобочитаемым (просмотрите изображение в окне **Крупный план**).

яркость – пример хорошего (в смысле распознавания) изображения.

Если в полученном изображении Вы обнаружили множество дефектов (разрывов или склеек букв), то обратитесь к нижеприведенной таблице, которая указывает возможные способы устранения этих дефектов.

| Особенности входного изображения: | Что сделать: |
|---|---|
|  <p>“разорванные”; светлые, тонкие буквы</p> | <ul style="list-style-type: none"> ● Уменьшить яркость (чтобы изображение стало темнее) ● Попробовать отсканировать в сером (в этом случае осуществляется автоподбор яркости) |
|  <p>искаженные и залитые; склеенные символы; темные, толстые буквы</p> | <ul style="list-style-type: none"> ● Увеличить яркость (сделать изображение светлее) ● Попробовать отсканировать в сером (в этом случае осуществляется автоподбор яркости) |

Сканирование многостраничных документов

Для удобства сканирования большого количества страниц в программе FineReader предусмотрен специальный режим сканирования: **Сканировать несколько страниц**.

Он позволяет отсканировать несколько страниц в цикле. При этом:

- при сканировании через TWAIN с использованием **интерфейса FineReader** по окончании сканирования страницы сканер автоматически начинает сканирование следующей;
- при сканировании через TWAIN с использованием **интерфейса TWAIN-драйвера** сканера **Twain-диалог** сканера не закрывается после окончания сканирования первой страницы. Вы можете положить следующую страницу в сканер и отсканировать ее. И т.д.

Вы можете отсканировать большое количество страниц двумя путями: с использованием автоподатчика (ADF) и без него.

При использовании автоподатчика (ADF)

- **Интерфейс FineReader.** В диалоге Настройки сканера (меню **Сервис>Настройки сканера**) отметьте опцию **Использовать автоподатчик бумаги (ADF)** и запустите сканирование нескольких страниц (меню **Файл>Сканировать несколько страниц...**).
- **Интерфейс TWAIN-драйвера сканера.** В TWAIN-диалоге сканера отметьте опцию **Use ADF** (может называться по-разному в зависимости от модели сканера; смысл опций описан в документации, прилагаемой к Вашему сканеру) и запустите сканирование (меню **Файл>Сканировать несколько страниц...**).

Без использования автоподатчика (ADF)

1. Интерфейс FineReader

- В меню **Файл** выберите пункт **Сканировать несколько страниц**.

Для удобства сканирования нескольких страниц подряд на планшетном сканере без автоподатчика:

- Установите величину паузы (время от окончания сканирования одной страницы до начала сканирования следующей). Для этого в диалоге Настройки сканера (меню **Сервис>Настройки сканера...**) отметьте опцию **Пауза между страницами** и установите величину паузы в секундах.

В этом случае после сканирования первой страницы сканер делает указанную паузу, во время которой Вы вставляете в сканер следующую страницу. Затем сканирование автоматически продолжается.

- Выберите режим **Останавливаться между страницами**. Для этого в диалоге Настройки сканера (меню **Сервис>Настройки сканера...**) отметьте опцию **Останавливаться между страницами**.

В этом случае каждый раз по окончании сканирования страницы будет появляться диалог с вопросом о продолжении сканирования. Нажмите кнопку **Да** для сканирования следующей страницы и **Нет** – для завершения сканирования.

Для завершения сканирования в меню **Файл** выберите пункт **Остановить сканирование**.

2. Интерфейс TWAIN-драйвера сканера

- В меню **Файл** выберите пункт **Сканировать несколько страниц**. Для начала сканирования в открытом TWAIN-диалоге сканера нажмите кнопку **Scan (Final)** и т.д., название кнопки зависит от конкретной реализации TWAIN-драйвера сканера).

Для продолжения сканирования нажмите на кнопку **Scan** в **Twain-диалоге** сканера.

Для завершения сканирования в **Twain-диалоге** сканера нажмите на кнопку **Close** (или аналогичную ей по смыслу).

Совет: Чтобы Вы могли контролировать результаты сканирования, на закладке Сканирование/Открытие (меню **Сервис>Опции**) отметьте опцию **Открывать изображения по мере сканирования**. В этом случае по окончании сканирования страницы отсканированное изображение появится в окне **Изображение**. Если изображение отсканировано неверно, остановите сканирование (в меню **Файл** выберите пункт **Остановить сканирование**) и отсканируйте изображение заново.

Открытие файлов с изображениями

Если у Вас нет сканера, Вы можете распознавать графические файлы (список открываемых форматов приведен в разделе “Поддерживаемые графические форматы”).

Чтобы открыть изображение:

- Нажмите стрелку справа от кнопки **1–Сканировать** и в локальном меню выберите пункт **Открыть изображение**. Внешний вид значка изменится; подпись **Сканировать** поменяется на **Открыть**.
- В меню **Файл** выберите пункт **Открыть изображение**.
- В Windows Explorer: щелкните правой кнопкой мыши на файле с изображением и в локальном меню выберите пункт **Открыть с помощью FineReader**. Если на Вашем компьютере уже открыт FineReader, изображение будет добавлено в текущий пакет; в противном случае перед добавлением изображения в пакет автоматически запустится FineReader с пакетом, с которым Вы работали в последний раз.
- В Microsoft Outlook и/или Windows Explorer: Нажмите левой кнопкой мыши на файле с изображением, которое Вы хотите открыть, и не отпуская кнопки, перетащите его на свернутое окно программы FineReader. Изображение будет добавлено в текущий пакет и открыто в окне **Изображение**.

В диалоге **Открыть (Open)** выберите одно или несколько изображений. Выбранные изображения появятся в окне **Пакет**, и последнее из выбранных изображений откроется в окне **Изображение** и в окне **Крупный план** на экране FineReader, при этом копия изображения помещается в папку пакета. Более подробно о представлении страниц в пакете и об устройстве пакета см. “Общая информация по работе с пакетом”.

Совет: Если Вы хотите, чтобы открытые изображения были сразу распознаны, воспользуйтесь режимом **Открыть и распознать**. Для этого:

1. В меню **Процесс** выберите пункт **Открыть и распознать**.
Клавиатурная команда: **Ctrl+Shift+D**.
 2. В открывшемся диалоге **Открыть (Open)** выберите изображения для распознавания.
-

Передача изображений через папку на локальном диске, в локальной сети или на FTP сервере



(только для версии Corporate Edition)

Если в Вашей компании используются многофункциональные сетевые устройства, объединяющие в себе функции сканирования, печати и копирования, Вы можете настроить FineReader так, чтобы изображения с такого устройства автоматически попадали в FineReader.

Одним из средств достижения этой цели служит режим мониторинга заданной папки на локальном диске, в локальной сети или на FTP сервере. В этом режиме программа автоматически открывает все добавленные в заданную папку отсканированные изображения, факсы, а также документы в формате PDF. После добавления в пакет FineReader исходное изображение стирается.

Чтобы активировать режим просмотра папки:

- В меню **Файл** выберите пункт **Просмотр папки...**, или
- В локальном меню кнопки **Сканировать** выберите пункт **Просмотр папки**.

При запуске режима мониторинга заданной папки меняется значок кнопки на панели **Scan&Read**, открывается диалог **Просмотр папки** и в строке состояния отображается значок . В случае возникновения сообщения об ошибке этот значок меняется на . Чтобы прочесть сообщение об ошибке, дважды щелкните мышкой на значке.

Чтобы отключить режим просмотра папки:

- В меню **Файл** выберите пункт **Отключить режим просмотра папки**, или
- На панели **Scan&Read** нажмите кнопку **Отключить**.

Добавление в пакет изображений со сдвоенными страницами

При сканировании книг удобнее отсканировать две страницы (книжный разворот) сразу. При этом для повышения качества распознавания такие изображения следует разделить на два, чтобы каждой странице соответствовала отдельная страница пакета (анализ и распознавание осуществляется для каждой страницы по отдельности, исправляется перекос строк).

Для этого перед сканированием или добавлением в пакет сдвоенных страниц:

- На закладке **Сканирование/Открытие** (меню **Сервис>Опции**) отметьте опцию **Делить разворот книги**.

В этом случае книжный разворот (сдвоенные страницы) будет представлен двумя страницами пакета. Подробнее о пакете см. “Общая информация по работе с пакетом”.

Замечание: Если книжный разворот (сдвоенные страницы) был разделен на две страницы неверно, то снимите отметку с опции **Делить разворот книги**, заново отсканируйте или добавьте в пакет изображение с книжным разворотом и попробуйте разделить его в диалоге **Разбить изображение** (меню **Изображение>Разбить изображение**) вручную.



Добавление в пакет изображений с визитными карточками

При вводе в компьютер информации с визитных карточек их удобнее сканировать не по одной, а сразу по несколько штук, сколько помещается на Ваш сканер. При этом качество распознавания будет выше (в частности, благодаря исправлению перекосов) в том случае, когда каждая визитка попадет в пакет в виде отдельной страницы. Для этого в системе предусмотрены средства автоматического и ручного разбиения изображений, содержащих расположенные в определенном порядке визитные карточки (см. “Обучение на примерах” программы ABBYY FineReader, раздел “Работа с визитными карточками”).

Чтобы разбить изображение:

1. В окне **Пакет** выберите нужное изображение.
2. В меню **Изображение** выберите пункт **Разбить изображение**.
3. В открывшемся диалоге **Разбить изображение** нажмите кнопку **Разбить на визитки**.

Замечания:

1. Разрезаемая страница удаляется из пакета; на её место добавляются новые страницы, соответствующие разрезанным частям. Подробнее о пакете см. “Общая информация по работе с пакетом”.
2. Если изображение было разрезано на визитки неверно, то попробуйте разрезать его вручную, воспользовавшись кнопками **Добавить вертикальный разделитель**/**Добавить горизонтальный разделитель**.
3. Чтобы удалить все разделители, нажмите кнопку **Удалить все разделители**.
4. Чтобы передвинуть разделители, перейдите в режим **Выбора объекта** (нажмите кнопку ) и переместите разделитель.
5. Чтобы удалить разделитель, перейдите в режим **Выбора объекта** (нажмите кнопку ) и переместите разделитель за пределы изображения.

Нумерация страниц при добавлении в пакет

По умолчанию каждой сканируемой странице присваивается номер на единицу больший номера последнего изображения в пакете.

Вы можете задать номер добавляемой страницы и вручную (например, Вам нужно сохранить исходную нумерацию страниц или Вы сканируете стопку сортированных по порядку страниц). Для этого:

- на закладке Сканирование/Открытие (меню **Сервис>Опции**) отметьте пункт **Запрашивать номер страниц при добавлении в пакет**.

При сканировании стопки двусторонних сортированных по порядку страниц:

1. Отметьте пункт **Запрашивать номер перед добавлением в пакет** на закладке Сканирование/Открытие (**Сервис>Опции**).
2. В диалоге **Номер страницы** укажите номер страницы, с которой начнется сканирование и выберите опцию **Через одну** в поле **Нумерация страниц**. Выберите способ нумерации страниц: по возрастанию или по убыванию. Возрастание или убывание зависит, например, от того, как Вы кладете стопку в автоподатчик – находятся ли меньшие или большие номера наверху.

Проверка и корректирование полученного изображения

- Очистить от мусора
- Инвертировать изображение
- Повернуть или зеркально отразить изображение
- Стереть участок изображения
- Увеличить/Уменьшить масштаб изображения
- Получить информацию об изображении
- Печать изображения
- Отменить последнее действие

1. Очистить от мусора

Распознаваемое изображение может быть сильно “замусорено”, т.е. содержать много лишних точек, возникших в результате сканирования документов среднего или плохого качества. Точки, близко расположенные к контурам букв, могут отрицательно сказаться на качестве распознанного текста. Чтобы уменьшить количество лишних точек, можно воспользоваться опцией **Очистить от мусора**. Для этого:

- В меню **Изображение** выберите пункт **Очистить изображение от мусора**.

Если Вы хотите очистить от “мусора” отдельный блок, то:

- В меню **Изображение** выберите пункт **Очистить блок от мусора**.

Внимание! Если исходный текст был очень светлым или в исходном тексте использовался очень тонкий шрифт, то применение функции **Очистить изображение от мусора** может привести к исчезновению точек, запятых или тонких элементов букв, что ухудшает качество распознавания.

Замечание: Если Вы сканируете или открываете “замусоренные” изображения, то перед добавлением в пакет таких изображений в группе **Обработка изображений** на закладке Сканирование/Открытие (меню **Сервис> Опции**) отметьте пункт **Очистить изображение от мусора**.

2. Инвертировать изображение

Некоторые сканеры инвертируют изображения при сканировании (черный цвет переводят в белый, а белый в черный). Чтобы получить стандартное представление документа (черный шрифт на белом фоне):



- В меню **Изображение** выберите пункт **Инвертировать**.

Замечание: Если Вы сканируете или открываете инвертированные изображения, то перед добавлением в пакет таких изображений в группе **Обработка изображений** на закладке Сканирование/Открытие (меню **Сервис> Опции**) отметьте пункт **Инвертировать**.

3. Повернуть или зеркально отразить изображение

При распознавании изображение должно иметь стандартную ориентацию: текст должен читаться сверху вниз, и строки должны быть горизонтальными. По умолчанию программа при распознавании определяет и корректирует ориентацию изображения автоматически. Если ориентация изображения была определена ошибочно, то на закладке Сканирование/Открытие снимите отметку с пункта **Определять ориентацию страницы (в процессе распознавания)** и поверните изображение вручную.

Чтобы повернуть изображение:


- на 90 градусов вправо – нажмите кнопку  или выберите в меню **Изображение** пункт **Повернуть по часовой стрелке**.
- на 90 градусов влево – нажмите кнопку  или выберите в меню **Изображение** пункт **Повернуть против часовой стрелки**.
- на 180 градусов – выберите в меню **Изображение** пункт **Повернуть на 180 градусов**.

Чтобы зеркально отразить изображение относительно:



- горизонтальной прямой – выберите в меню **Изображение** пункт **Зеркально отразить относительно горизонтали**;
- вертикальной прямой – выберите в меню **Изображение** пункт **Зеркально отразить относительно вертикали**.

4. Стереть участок изображения

Если Вы хотите исключить какой-то участок текста из распознавания или у Вас на изображении имеются большие участки мусора, то Вы можете стереть такие участки. Для этого:

- Выберите инструмент  (на панели в окне **Изображение**) и, нажав на левую кнопку мыши, выделите участок изображения, который Вы хотите удалить. Отпустите кнопку, выделенная часть изображения будет удалена.

5. Увеличить/Уменьшить масштаб изображения

- На панели **Изображение** (в окне **Изображение**) выберите инструмент  /  и щелкните мышью на изображении. Изображение увеличится/уменьшится в два раза.
- Щелкните правой кнопкой мыши на изображении и в локальном меню выберите пункт **Масштаб** и нужный Вам масштаб.

6. Получить информацию об изображении

Вы можете получить следующую информацию об открытом изображении:

ширину и высоту изображения в точках; вертикальное и горизонтальное разрешение в точках на дюйм (dpi); тип изображения.


- Щелкните правой кнопкой на изображении и в локальном меню выберите пункт **Свойства**. В открывшемся диалоге выберите закладку **Изображение**.


7. Печать изображения

Вы можете напечатать одно изображение, открытое в окне **Изображение**, несколько изображений, выделенных в окне **Пакет**, или все изображения. Для этого:

- В меню **Файл** выберите пункт **Печать изображения** и в открывшемся диалоге **Печать** установите параметры печати (принтер, количество печатаемых страниц, количество копий и тд.)

8. Отменить последнее действие

- Для отмены последнего действия на панели **Стандартная** нажмите кнопку **Отменить** .

Совет: Для повторного выполнения последнего отмененного действия на панели **Стандартная** нажмите кнопку **Вернуть** .

Опции хранения изображений в пакете

Приводить цветное/серое изображение к черно-белому (меню Сервис>Опции, закладка Сканирование/Открытие)

Отметьте эту опцию при сканировании через TWAIN-диалог сканера в сером режиме (с автоподбором яркости) или при сканировании в цвете, если при этом сканируемые документы не содержат цветных картинок, цветного шрифта и фона или же Вам не требуется передача цвета в выходное изображение. В этом случае сохраняемые в пакет изображения будут занимать меньше места на диске.

Глава 5

Анализ макета страницы

Прежде чем приступить к распознаванию, программа должна знать, какие участки изображения надо распознавать. Для этого проводится анализ макета страницы, во время которого выделяются блоки с текстом, картинки, таблицы и штрих-коды.

В этой главе Вы узнаете, когда может потребоваться провести ручной анализ макета страницы, какие типы блоков бывают, как можно отредактировать полученные в результате автоматического анализа блоки, а также, как можно упростить процесс анализа, используя шаблоны блоков.

Краткое содержание раздела:

- Общая информация по анализу макета страницы
- Типы блоков
- Опции автоматического анализа макета страницы
- Выделение и редактирование блоков вручную
- Ручной анализ таблицы
- Использование шаблонов блоков

Общая информация по анализу макета страницы

Анализ макета страницы может проводиться как автоматически, так и вручную. В большинстве случаев FineReader сам успешно справляется с анализом сложных страниц. Автоматический анализ производится по кнопке 2–Распознать одновременно с распознаванием текста.

Замечание: Отдельная процедура анализа макета страницы тоже доступна (меню **Процесс>Анализ макета страницы**). Правда, при этом качество сегментации может быть ниже, т.к. при совместной процедуре распознавания и сегментации для анализа страницы используется дополнительная информация, полученная в процессе распознавания.

Ручное выделение блоков может понадобиться, если

1. Вы хотите распознать **часть** страницы;
2. в результате автоматического анализа блоки были выделены неправильно.

Совет:

- В некоторых случаях качество автоматического анализа можно улучшить, правильно установив опции анализа макета. Проверьте установленные опции анализа (закладка Распознавание, меню **Сервис>Опции**).
 - В случае если программа выделила неправильно некоторые блоки, часто оказывается быстрее исправить только их, воспользовавшись инструментами для редактирования блоков, а не выделять блоки на изображении заново вручную.
-

Типы блоков

Блоки – это заключенные в рамку участки изображения. Блоки выделяют для того, чтобы указать системе, какие участки отсканированной страницы надо распознавать и в каком порядке. Также по ним воспроизводится исходное оформление страницы. Блоки разных типов имеют различные цвета рамок. Вы можете изменить цвета рамок блоков на закладке **Вид** диалога **Опции** (меню **Сервис>Опции**) в группе **Объекты**. В поле **Объект** выберите нужный тип блока, а в поле **Цвет** – требуемый цвет.

При обработке изображений выделяют блоки следующих типов:

Зона Распознавания – блок используется для распознавания и автоматического анализа части изображения. После нажатия на кнопку **2–Распознать** выделенный блок автоматически анализируется и распознается.

Текст – блок используется для обозначения текста. Он должен содержать только одноколоночный текст. Если внутри текста содержатся картинки, выделите их в отдельные блоки.

Таблица – этот блок используется для обозначения таблиц или текста, имеющего табличную структуру. При распознавании программа разбивает данный блок на строки и столбцы и формирует табличную структуру. В выходном тексте данный блок передается таблицей.

Вы можете выделить и отредактировать таблицу вручную.

Картинка – этот блок используется для обозначения картинок. Он может содержать картинку или любую другую часть текста, которую Вы хотите передать в распознанный текст в качестве картинки.

Штрих-код – этот блок используется для распознавания штрих-кодов. Т.е., если Ваш документ содержит штрих-код и Вы хотите передать его не картинкой, а перевести его в последовательность букв и цифр, то выделите штрих-код в отдельный блок и присвойте ему тип **Штрих-код**.

Замечание. По умолчанию опция, позволяющая искать и распознавать штрих-коды отключена. Чтобы подключить ее, отметьте пункт **Искать штрих-коды** на закладке **Распознавание** (меню **Сервис>Опции**).

Опции автоматического анализа макета страницы

При автоматическом анализе макета страницы FineReader сам выделяет блоки, содержащие тексты, таблицы, картинки и штрих-коды.

Автоматический анализ запускается по кнопке **2–Распознать** одновременно с распознаванием текста. До запуска распознавания необходимо установить основные опции анализа: тип страницы и опции анализа таблиц.

Тип страницы

Для большинства изображений расположение текста на странице определяется автоматически, чему соответствует значение **Авто** на закладке **Распознавание** в группе **Тип страницы** (меню **Сервис>Опции**), устанавливаемое системой по умолчанию.

В некоторых случаях может потребоваться установить значение типа страницы вручную. Для этого:

- На закладке **Распознавание** диалога **Опции** (меню **Сервис>Опции**) в группе **Тип страницы** выберите нужный Вам пункт.

Возможные типы страницы:

Автоматическое определение – указывает, что расположение текста на странице определяется автоматически. Это значение устанавливается системой по умолчанию; подходит для распознавания всех видов текстов, в том числе многоколоночного текста, текста с таблицами и картинками.

Одна колонка – указывает, что текст на странице напечатан в одну колонку. Эта опция используется в случае, если автоматическое определение ошибочно сегментировало страницу как многоколоночный текст.

Форматированный пробелами текст – указывает, что текст на странице расположен в одну колонку и напечатан моноширинным шрифтом одного размера. В распознанном тексте сохраняется деление на строки; отступы от левого края передаются пробелами; каждая строка выделяется в отдельный абзац, и расстояния между абзацами передаются пустыми строками. Используется, например, для распознавания распечаток текстов программ.

Опции анализа таблиц

В большинстве случаев программа делит таблицу на строки и столбцы автоматически. Дополнительная настройка опций анализа таблиц устанавливается на закладке **Распознавание** в группе **Таблицы**. Эти опции рекомендуется использовать, если:

- в результате автоматического анализа макета страницы таблица была выделена и разделена на строки и столбцы неверно;
- документ содержит много однотипных таблиц, для которых известна дополнительная информация (например: таблица не содержит объединенных ячеек или таблица состоит из ячеек, текст в которых расположен в одну строку).

1. Опция **В каждой ячейке таблицы не более одной строки текста** используется для уверенного анализа таблиц с неполным количеством или без черных разделителей, с ячейками, содержащими не более одной строки текста.

| Kilometers | Miles |
|------------|-------|
| 1 | 0.62 |
| 5 | 3.2 |

– пример таблицы, в каждой ячейке которой текст расположен в одну строку

| Физ. явление | t по Цельсию |
|-----------------------------|--------------|
| Температура кипения воды | 100 |
| Температура замерзания воды | 0 |

– пример таблицы, ячейки которой содержат более одной строки текста

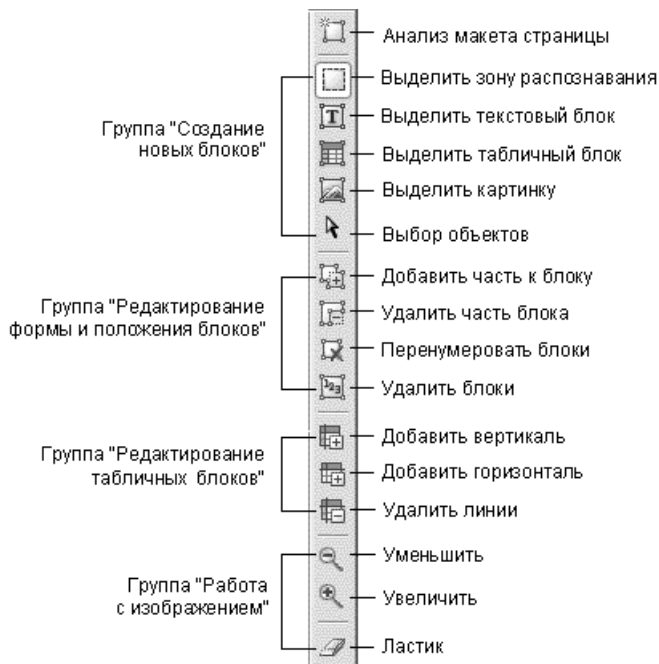
2. Опция **Таблица не содержит объединенные ячейки** используется для уверенного анализа таблиц, не содержащих объединенных ячеек.

| Температура | |
|-------------|-------------|
| По Цельсию | По Кельвину |
| -273 | 0 |
| 100 | 373 |

– пример таблицы с объединенной ячейкой

Замечание: На таблицах общей структуры или на страницах с таблицами различной структуры отмеченные опции **В каждой ячейке таблицы не более одной строки текста** и **Таблица не содержит объединенных ячеек** могут привести к ошибкам анализа и соответственно снизить качество распознавания.

Выделение и редактирование блоков вручную







Чтобы:

- Создать новый блок
- Отредактировать форму и расположение блоков
- Отредактировать таблицу

Чтобы создать новый блок:

1. Выберите один из инструментов:

-  – выделить зону распознавания;
-  – выделить текстовый блок;
-  – выделить картинку;
-  – выделить табличный блок.

2. Установите курсор мыши в угол предполагаемого блока. Нажмите левую кнопку мыши и, не отпуская кнопки, потяните в противоположный по диагонали угол.
3. Отпустите кнопку мыши.

Выделенная часть изображения будет заключена в рамку.

Вы можете поменять тип блока (присвоить выделенному блоку один из существующих типов: Зона распознавания, Текст, Таблица, Картинка или Штрих-код). Для этого:

- Щелкните на блоке правой кнопкой мыши и в локальном меню выберите **Тип блока**, а затем – нужный Вам пункт.


Редактирование формы и положения блоков

Чтобы передвинуть границу блока:


1. Установите курсор мыши на границу блока.
2. Нажмите левую кнопку мыши и потяните в нужную сторону.
3. Отпустите кнопку мыши.

Замечание: Если Вы установите курсор мыши на угол блока, то при движении мыши будут одновременно изменяться вертикальная и горизонтальная границы блока.

Чтобы добавить прямоугольную часть блока:

1. Выберите инструмент .
2. Установите курсор мыши внутри блока, к которому Вы хотите добавить часть. Нажмите левую кнопку мыши и, не отпуская кнопки, потяните по диагонали. Выделив нужную часть изображения, отпустите кнопку мыши. Выделенный прямоугольник будет добавлен к блоку.
3. Если нужно, передвиньте границу блока.

Чтобы удалить прямоугольную часть блока:


1. Выберите инструмент .
2. Установите курсор мыши внутри блока, там, где Вы хотите вырезать часть. Нажмите левую кнопку мыши и, не отпуская кнопки, потяните по диагонали. Выделив нужную часть изображения, отпустите кнопку мыши. Выделенный прямоугольник будет удален из блока.
3. Если нужно, передвиньте границу блока.









Замечания:

1. Вы также можете изменять границы блоков, добавляя на них новые узлы (точки разбиения). Получившиеся отрезки можно перемещать мышью в любом направлении.
Чтобы добавить новый узел, подведите курсор, удерживая клавишу Shift, к нужной точке границы (курсор при этом примет форму перекрестия) и щелкните мышью. На границе блока появится новый узел.






2. FineReader накладывает на допустимую форму блоков некоторые ограничения, обусловленные необходимостью передавать на распознавание неразрывные текстовые строки. Для соблюдения этих ограничений FineReader либо автоматически корректирует границы блока после операций добавления или удаления прямоугольной части блока (например, при удалении внутренней части блока снизу или сверху дополнительно удаляется часть блока справа до границы блока), либо запрещает некоторые операции по передвижению отрезков, образующих границу блока.

Чтобы выделить один или несколько блоков:


- Выберите инструмент  и щелкните мышкой по нужному блоку или, удерживая кнопку мыши, нарисуйте прямоугольник, охватывающий нужные блоки.

Замечание: Вы можете выделить один или несколько блоков, используя стандартные инструменты выделения блоков. Чтобы выбрать несколько блоков, нажмите клавишу **Shift** или **Ctrl** (при этом должен быть выбран один из инструментов: , ,  или ) и мышью щелкните на требуемых блоках. Чтобы отменить выделение уже выбранного блока или добавить невыделенные блоки, нажмите клавишу **Ctrl** (при этом должен быть выбран один из инструментов: , ,  или ) и мышью щелкните на требуемых блоках.

Чтобы передвинуть блок:


- Нажмите клавишу **Alt** (при этом должен быть выбран один из инструментов: , , ,  или ) и мышью переместите блоки.

Чтобы перенумеровать блоки:

1. Выберите инструмент .
2. Выделите блоки в том порядке, в котором Вы хотите видеть их содержимое в выходном тексте.

Замечание: Если Вы перенумеровываете блоки на уже распознанном изображении, то одновременно в окне **Текст** в черновом режиме редактора происходит перегруппировка распознанного текста в соответствии с новой нумерацией.

Чтобы удалить блок:

- Выберите инструмент  и выделите блок, который Вы хотите удалить.
- Выделите блоки, которые Вы хотите удалить, и нажмите клавишу **DEL**.

Внимание! Если Вы удаляете блок с уже распознанного изображения, то одновременно с этим в окне **Текст** удаляется текст, соответствующий этому блоку.

Чтобы удалить все блоки на изображении:

- В меню **Пакет** выберите пункт **Удалить блоки и текст**.




Внимание! Если Вы удаляете блоки с уже распознанного изображения, то одновременно с этим в окне **Текст** удаляется текст, соответствующий этим блокам.

Ручной анализ таблицы

Совет: Если в результате автоматического анализа таблицы разделение на строки и столбцы произошло неверно, прежде чем анализировать таблицу вручную заново, попробуйте сначала отредактировать результаты автоматического анализа.

Чтобы отредактировать таблицу вручную:

Отредактируйте таблицу, используя инструменты (панель в окне **Изображение**):

-  – Добавить вертикальную линию.
-  – Добавить горизонтальную линию.
-  – Удалить линию.

Если ячейка таблицы содержит только картинку, в диалоге **Свойства блока** (меню **Вид>Свойства**) отметьте пункт **Считать ячейку картинкой**. Если же, помимо картинки, в ячейке содержится некоторый текст, то выделите картинку в отдельный блок внутри ячейки.

Чтобы объединить ячейки или строки таблицы:

- в меню **Правка** выберите пункты **Объединить ячейки таблицы** или **Объединить строки таблицы**. При объединении строк таблицы деление на столбцы остается.

Объединенные ячейки можно снова разделить, воспользовавшись командой **Разбить ячейки таблицы** (меню **Правка**).

Замечание: Чтобы повторно не рисовать вертикальные и горизонтальные линии в таблице, выделите таблицу в отдельный блок и щелкните правой кнопкой мыши на блоке. В локальном меню выберите пункт **Анализ структуры таблицы**. А затем, используя инструменты по работе с таблицей, отредактируйте полученные результаты.

Использование шаблонов блоков

При работе с документами с одинаковым расположением текста и картинок, такими, например, как формы, бланки и т.п., вместо того чтобы анализировать макет каждой страницы, Вы можете провести анализ одной из них, сохранить расположение блоков на этой странице в файл, а затем, когда потребуется, “спроецировать” эти блоки на изображение (или группу изображений) со сходным расположением текста.

Замечание: Созданный шаблон блоков можно использовать только для документов, отсканированных с тем же разрешением, что и документ, на котором данный шаблон блоков создавался.

Чтобы создать шаблон блоков:

1. Откройте изображение и выделите на нем блоки автоматически или вручную.
2. В меню **Изображение** выберите пункт **Сохранить блоки...** В открывшемся диалоге укажите имя для шаблона блоков.

Чтобы наложить шаблон блоков:

1. В окне **Пакет** выделите страницы, на которые Вы хотите наложить существующий шаблон.
2. В меню **Изображение** выберите пункт **Наложить блоки**. В открывшемся диалоге **Открыть файл с блоками** выберите файл (*.blk) с нужным Вам расположением блоков.
3. В диалоге **Открыть файл с блоками** в группе **Применить к** установите переключатель в одно из положений: **Всем страницам** (если Вы хотите наложить шаблон на все страницы пакета) или **Выделенным страницам** (если Вы хотите наложить шаблон только на выделенные страницы).
4. Нажмите кнопку **Открыть (Open)**.

Глава 6

Распознавание

Задача распознавания состоит в том, чтобы преобразовать отсканированное изображение в текст, сохранив при этом оформление страницы. Прежде чем приступить к распознаванию текста, необходимо установить основные параметры распознавания: язык распознавания, тип печати распознанного текста и тип страницы. В этой главе описываются эти и другие параметры распознавания и приводятся ситуации, в которых они используются.

Краткое содержание раздела:

- Общая информация по распознаванию
- Язык распознавания
- Тип печати входного текста
- Другие опции распознавания
- Распознавание в фоновом режиме
- Распознавание с обучением
- Обучение эталона
- Редактирование эталона
- Пользовательские языки и группы языков
- Создание нового языка
- Создание новой группы языков

Общая информация по распознаванию

Внимание! Перед запуском распознавания проверьте установленные опции: язык распознавания, тип печати распознаваемого текста и тип страницы.

Вы можете:

1. Распознать блок или несколько блоков, выделенных на изображении.
2. Распознать открытую страницу или все страницы, выделенные в окне **Пакет**.
3. Распознать все нераспознанные страницы пакета.
4. Распознать все страницы в фоновом режиме. В этом режиме возможно распознавание с одновременным редактированием уже распознанных страниц.
5. Распознать страницы в режиме распознавание с обучением. Данный режим применяется в основном для распознавания текстов, использующих декоративные шрифты, или для распознавания большого объема (более 100 страниц) документов плохого качества печати.
6. Распознать страницы одного пакета на нескольких компьютерах одновременно.

Чтобы запустить распознавание:

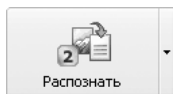
- Нажмите кнопку **2-Распознать** на панели **Scan&Read**.
- В меню **Процесс** выберите нужный Вам пункт:

Распознать – чтобы распознать открытую страницу или все страницы, выделенные в окне **Пакет**;

Распознать все – чтобы распознать все нераспознанные страницы пакета;

Распознать Блок – чтобы распознать блок или несколько блоков, выделенные на изображении;

Запустить фоновое распознавание – чтобы запустить распознавание в фоновом режиме.



Кнопка **2-Распознать** запускает распознавание открытого изображения. Чтобы изменить режим кнопки, нажмите на стрелку справа от нее и из открывшегося меню выберите нужный пункт.

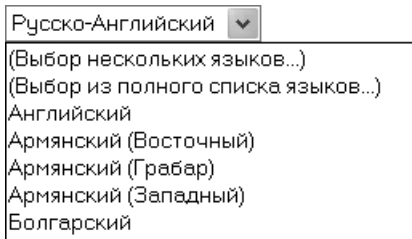
Замечание: При распознавании уже распознанной страницы перераспознаются только отредактированные и добавленные блоки.

Язык распознавания

FineReader поддерживает распознавание как одноязычных, так и многоязычных (например, английско–французских) документов. При распознавании текстов на английском или немецком языках Вы можете воспользоваться одноименными языками с медицинскими и юридическими словарями. Данные языки, помимо тематических словарей, включают в себя также и обычные словари FineReader.

Чтобы указать язык распознаваемого текста:

- Выберите соответствующую строку в списке на панели **Распознавание**.



Если Вы хотите распознать документ, написанный на нескольких языках:

1. В списке языков на панели **Стандартная** выберите пункт **Выбор нескольких языков...**
2. В открывшемся диалоге **Язык распознаваемого текста** укажите несколько языков. Для этого:
 - отметьте пункты с соответствующими названиями языков.

Совет: Если Вы часто используете какую–либо комбинацию языков, то создайте новую группу, содержащую эти языки.

Внимание!

1. Увеличение количества подключенных к распознаванию одного документа языков может привести к ухудшению качества распознавания. Не рекомендуется подключать более 2–3 языков.
2. Перед запуском распознавания проверьте подключенные на закладке **Форматирование** шрифты: они должны содержать все символы языка распознавания. В противном случае распознанный текст будет неправильно отображен в окне **Текст** (в словах на месте некоторых букв стоят значки “?” или “□”).

Если нужного языка нет в списке, возможны следующие варианты:

1. Данный язык не поддерживается системой FineReader. Полный список языков приведен в разделе “Поддерживаемые языки распознавания” справки программы ABBYY FineReader.
2. Язык исключен из списка языков показываемых на панели **Стандартная**. В этом случае в списке языков на панели **Стандартная** выберите пункт **Выбор из полного списка языков** и в открывшемся диалоге **Язык распознаваемого текста** укажите необходимый язык.
3. Язык был отключен при выборочной установке. Чтобы доустановить языки распознавания, запустите программу инсталляции FineReader в режиме покомпонентной установки (**Установка дистрибутива по выбору**), снимите отметки со всех пунктов, кроме **Языки распознавания**, и нажмите кнопку **Состав**. В открывшемся списке языков укажите требуемые языки.

Замечание: При установке проверьте, что Вы указали ту же папку, в которую Вы ранее установили ABBYY FineReader.

Чтобы подключить/отключить язык к списку показываемых языков:

В диалоге Редактор языков (меню **Сервис>Редактор языков**) выберите язык, который Вы хотите подключить/отключить, и отметьте пункт (снимите отметку) **Показывать в списке языков**.

Совет: Вы можете установить язык распознавания на отдельный блок. Для этого щелкните правой кнопкой мыши на блоке, для которого вы хотите установить язык распознавания, отличный от языка распознавания для всего текста, и из локального меню выберите пункт **Свойства**. В открывшемся диалоге **Свойства** на закладке **Блок** в поле **Язык распознавания** выберите язык распознавания выделенного блока.

Тип печати входного текста

Для большинства текстов тип печати определяется автоматически. Этому соответствует значение **Авто** (группа **Тип печати**, меню **Сервис>Опции**, закладка **Распознавание**).

При распознавании текстов, напечатанных на матричном принтере в черновом режиме или на пишущей машинке, можно добиться более высокого качества распознавания, установив правильный **Тип печати**:

- для текстов, напечатанных на пишущей машинке – значение **Пишущая машинка**;
- для текстов, напечатанных на матричном принтере – значение **Матричный принтер**.

software

Фрагмент страницы, напечатанной на матричном принтере в черновом режиме. На картинке видно, что штрихи букв состоят из отдельно стоящих точек.

software

Фрагмент страницы, напечатанной на пишущей машинке. Ширина букв одинакова (сравните, например, буквы “w” и “t”).

Чтобы поменять тип печати:

- На закладке **Распознавание** диалога **Опции** (меню **Сервис>Опции**) в группе **Тип печати** выберите требуемый пункт.

Замечание: После распознавания текстов, напечатанных на пишущей машинке или матричном принтере, не забудьте снова выбрать значение **Авто** при возвращении к типографскому тексту.

Другие опции распознавания

Показывать или не показывать изображение при распознавании

Распознавание группы страниц происходит быстрее, если обрабатываемое изображение не показывается.

- На закладке **Опции** (меню **Сервис>Опции**, диалог **Опции**) снимите отметку с пункта **Показывать изображение при распознавании**.

Указать ориентацию текста

Если программа некорректно распознала блоки с вертикальным направлением текста (текстовый блок, ячейки таблицы):

- Щелкните правой кнопкой мыши по блоку с вертикальным направлением текста и в локальном меню выберите пункт **Свойства**. В открывшемся диалоге **Свойства** выберите нужный вариант в списке **Направление текста** и распознайте изображение заново.



Указать, что текст в блоке является инвертированным или отраженным

Если программа некорректно распознала блоки с инвертированным или отраженным текстом (текстовый блок, ячейка таблицы или вся таблица), то:

- Щелкните правой кнопкой мыши на нужном блоке и в локальном меню выберите пункт **Свойства**. В открывшемся диалоге Свойства отметьте пункт **Инвертированный** или **Отраженный** соответственно и распознайте изображение заново.

Распознавание в фоновом режиме

Если Вы хотите одновременно с распознаванием редактировать распознанные страницы, Вы можете запустить распознавание в фоновом режиме:

- В меню **Процесс** выберите пункт **Запустить фоновое распознавание**. В строке состояния внизу экрана появится значок . Если для окна Пакег выбран режим показа свойства страниц (локальное меню, пункт **Вид>Вид со свойствами**), то напротив распознаваемой страницы в колонке **Открыта** появится значок .

В этом режиме распознавание автоматически возобновляется, как только в пакете появляются нераспознанные страницы.

Замечание: На многопроцессорных компьютерах режим “Фоновое распознавание” позволяет увеличить скорость распознавания пакетов, содержащих большое количество страниц.

Чтобы остановить Распознавание в фоновом режиме:

- В меню **Процесс** выберите пункт **Остановить фоновое распознавание**.

Внимание! В режиме работы **Распознавание в фоновом режиме** используются опции, установленные в программе до запуска фонового распознавания.

Распознавание с обучением

Как было отмечено, программа FineReader позволяет вводить тексты разного качества, напечатанные практически любыми шрифтами. Тексты хорошего и среднего качества, а также шрифты обычного начертания распознаются без предварительного обучения.

Режим “Распознавание с обучением” используется для:

1. распознавания текстов, использующих декоративные шрифты;
2. распознавания текстов, в которых встречаются специальные символы (например, отдельные математические символы);
3. распознавания большого объема (более 100 страниц) текста плохого качества.

В других случаях **Распознавание с обучением** использовать *не рекомендуется*, т.к. затраты на обучение будут больше, чем полученный выигрыш в качестве распознавания.

Обучение проводится при распознавании одной–двух страниц текста в специальном режиме. В результате создается эталон букв, встречающихся в тексте. Этот эталон в дальнейшем используется при распознавании основного объема текста. Некоторые пары или тройки символов в тексте могут склеиваться. Если при обучении Вам не удастся переместить описывающий прямоугольник так, чтобы он заключал в себя один целый символ и не содержал при этом части соседних, то Вы можете обучить программу сочетание символов, которые невозможно “расклеить”. Такие неразделяемые сочетания двух или трех символов называются лигатурами. Это, например, такие сочетания, как ed, ot, ff, ffi, ffl и другие.

Внимание!

1. Созданный эталон можно использовать только для распознавания текстов, использующих тот же шрифт и размер шрифта и отсканированных с тем же разрешением, что и документ, на котором данный эталон создавался.
2. При удалении пакета эталон также удаляется. Вы можете сохранить созданный эталон для работы с другими пакетами. Для этого сохраните настройки пакета в формате шаблон пакета (*.fnt).
3. При переходе к распознаванию текстов, напечатанных другим шрифтом, не забудьте отключить эталон (на закладке **Распознавание**, меню **Сервис>Опции** установите переключатель в положение **Не использовать пользовательский эталон**).

Последовательность действий при распознавании с обучением:

1. Установите режим **Распознавание с обучением** (на закладке **Распознавание**, меню **Сервис>Опции** в группе **Распознавание с обучением** установите переключатель в положение **Распознавание с обучением**). В строке состояния появится название эталона (по умолчанию default).
2. Нажмите кнопку **2–Распознать**.
3. Обучите эталон, распознав одну–две страницы в режиме распознавание с обучением.

Обучаемые символы заносятся в эталон, создаваемый системой по умолчанию. По окончании обучения система сохранит созданный эталон (default.ptn) в папке, где хранится пакет.

4. Отредактируйте эталон.
5. Отмените режим **Распознавание с обучением** (на закладке **Распознавание** в группе **Обучение** установите переключатель **Распознавание с пользовательским эталоном**).
6. Запустите распознавание основного текста, нажав на кнопку **2-Распознать**.

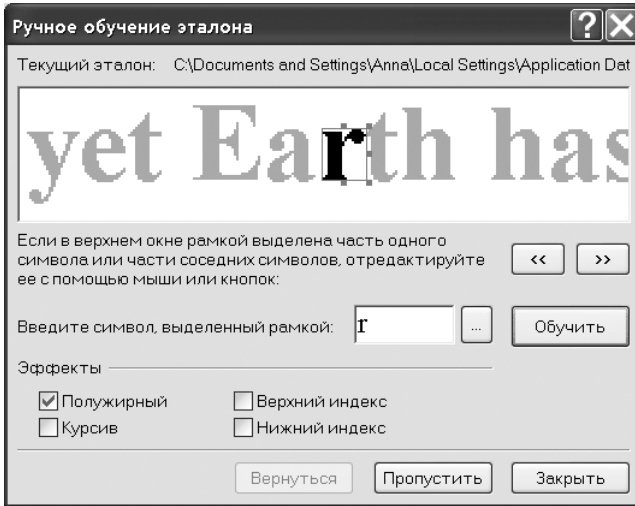
Замечания:

1. Чтобы создать несколько эталонов на один пакет, воспользуйтесь диалогом Редактор эталонов (вызывается с закладки **Распознавание** по кнопке **Эталоны** или из меню **Сервис>Редактор эталонов**). Создайте новый эталон (нажмите в диалоге на кнопку **Новый**) и выберите его для работы (нажмите на кнопку **Выбрать**). Далее работа с созданным эталоном происходит так же, как и работа с default-эталомом, см. выше пп 1–5.
2. Если в процессе обучения было создано несколько эталонов, то подключается последний созданный эталон. Название подключенного эталона пишется в строке состояния.
Чтобы подключить другой эталон для распознавания, в диалоге Редактор эталонов (меню **Сервис>Редактор эталонов**) в списке эталонов выберите эталон и нажмите кнопку **Выбрать**. На закладке **Распознавание** в группе **Обучение** установите переключатель в положение **Распознавание с пользовательским эталоном**.
3. Если на закладке **Распознавание** отмечена опция **Использовать встроенные эталоны**, то в режиме **Распознавание с обучением** программа предложит обучить только неуверенно распознанные символы.



Если Вы обучаете программу декоративным или нестандартным шрифтам (например, тибетскому) и при этом используете встроенные эталоны, программа может распознать часть символов неправильно, не предложив их обучить. В этом случае рекомендуется снять отметку с опции **Использовать встроенные эталоны**: обучение будет производиться каждому символу.

Обучение эталона

1. Проверьте, чтобы на закладке **Распознавание** (меню **Сервис>Опции**) в группе **Распознавание с обучением** переключатель был установлен в положение **Распознавание с обучением**.
2. Нажмите кнопку **2-Распознать**. Программа начнет распознавание. Как только встретится символ, подлежащий обучению, откроется диалог Ручное обучение эталона с изображением этого символа.




Обучение символу

Описывающий прямоугольник в верхней части диалога должен содержать **один целый символ**. Если он содержит часть буквы или более одной буквы, то с помощью мыши или кнопок  и  прямоугольник можно передвинуть так, чтобы он охватывал одну целую букву.

Далее введите нужный символ и нажмите кнопку **Обучить**.



Внимание!

1. Обучать можно только символам, входящим в алфавит языка. Если Вы обучаете программу символам, которые нельзя ввести с клавиатуры, то для их обозначения можно использовать комбинацию из двух символов или Вы можете скопировать требуемый символ из **Таблицы символов** (открывается при нажатии в диалоге **Ручное обучение эталона** кнопки ).
2. Если в обучаемом тексте встречаются слова, набранные *курсивом* или **полужирным**, и Вам важно сохранить гарнитуру шрифта в распознанном тексте, то при обучении таким символам в диалоге **Ручное обучение эталона** следует отметить пункты **Курсив** или **Полужирный** соответственно.
3. При обучении следите за тем, чтобы изображениям заглавных букв соответствовали заглавные буквы, а изображениям строчных букв – строчные.

Если при обучении Вы ошиблись, то можно нажать кнопку **Вернуться**, и охватывающий прямоугольник вернется на предыдущую позицию, а последняя обученная пара “изображение – символ” будет удалена из эталона. Кнопка **Вернуться** действует в пределах одного слова.

Обучение лигатурам

Лигатуры – это сочетания двух или трех символов, которые из-за особенностей их начертания невозможно разделить при обучении и которым поэтому сразу обучаются как комбинации символов. Обучение лигатурам происходит так же, как и обучение отдельным символам:

1. В строке для ввода символа введите нужное сочетание символов и нажмите кнопку **Обучить**.
2. Описывающий прямоугольник в верхней части диалога должен содержать **сочетание целиком**. Передвинуть прямоугольник можно с помощью мыши или используя кнопки  и .

В одном эталоне может содержаться до 1000 новых символов. Однако не следует создавать слишком много лигатур, т.к. это может отрицательно сказаться на качестве распознавания.

При обучении необходимо учесть следующие ограничения:

1. Изображения некоторых символов не различаются системой распознавания и сопоставляются с каким-то одним символом. Например, прямой (’), левый (‘) и правый (’) апострофы хранятся в эталоне как изображение прямого апострофа. Таким образом, в результате распознавания в тексте никогда не появится правый или левый апостроф, хотя при обучении Вы указывали именно эти символы.
2. Для некоторых изображений решение о том, какому символу в распознанном тексте сопоставить встретившееся конкретное изображение, принимается на основе общего анализа распознанного текста. Так, например, решение о том, является ли символ, обозначаемый “кружком”, буквой о или цифрой ноль, система принимает в зависимости от того, находятся ли рядом другие цифры или буквы.

Редактирование эталона

Прежде чем запускать распознавание с только что созданным эталоном, рекомендуется просмотреть эталон и, если потребуется, отредактировать. Этим Вы сведете к минимуму ошибки распознавания, которые могут возникнуть из-за неправильно обученного эталона.

Эталон должен содержать только целые символы или лигатуры. Символы, обрезанные с краев, и символы с неправильными подписями следует удалить из эталона.

Чтобы отредактировать эталон:

1. В меню **Сервис** выберите пункт **Редактор эталонов...**
2. В открывшемся диалоге **Редактор эталонов** выберите нужный эталон и нажмите кнопку **Редактировать...** Перед Вами откроется диалог **Символы пользовательского эталона**.
3. Выбрав символ, нажмите кнопку **Свойства**, чтобы отредактировать подпись и указать правильное начертание: курсив, полужирный, верхний или нижний индексы, или нажмите кнопку **Удалить**, чтобы удалить неправильно обученные символы.

Пользовательские языки и группы языков

Вы можете использовать не только predetermined языки и группы языков, но и создать новый язык или объединить существующие языки в новую группу и при распознавании подключить именно их.

Когда надо создавать новый язык:

1. Для подключения пользовательского словаря.
Например, необходимо распознать русский текст, содержащий аббревиатуры. Вы можете создать словарь аббревиатур и подключить его к пользовательскому языку. На основе русского языка с подключенным системным словарем и языка, созданного Вами с подключенным словарем аббревиатур, Вы можете создать группу для дальнейшего ее использования при распознавании Ваших текстов.
2. Для распознавания документов специального вида.
Например:
 - Страница содержит перечень артикулов, состоящий из цифр и нескольких букв. Вы можете создать новый язык, включив в него минимально необходимый набор символов, и использовать его для распознавания данного типа документов.
 - Документ использует только заглавные буквы английского языка. В этом случае для повышения качества распознавания следует исключить из распознавания символы, которые заведомо не могут встретиться в тексте, в данном случае – все строчные буквы.

Когда надо создавать группу языков:

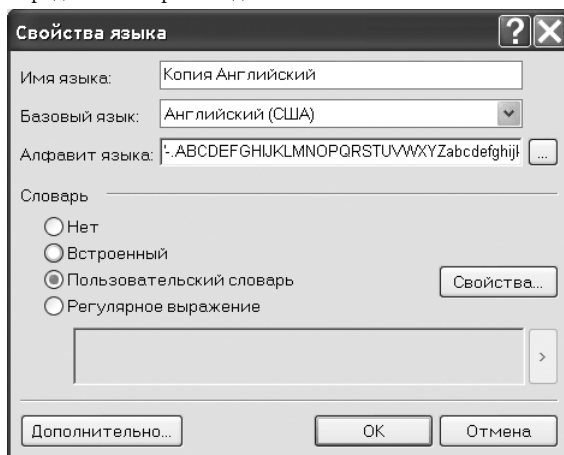
- Если Вы часто используете какую-нибудь комбинацию языков.

Создать язык или группу языков можно из диалога Редактор языков (меню **Сервис**, пункт **Редактор языков**).

Создание нового языка

Чтобы создать новый язык распознавания:

1. В меню **Сервис** выберите пункт **Редактор языков...**
2. Нажмите кнопку **Новый**. В открывшемся диалоге установите переключатель в положение **Создать новый язык на основе существующего** и выберите язык, на основе которого Вы создаете новый.
3. Перед Вами откроется диалог Свойства языка.



При создании нового языка нужно определить следующие параметры (все параметры задаются в диалоге Свойства языка):

1. Имя нового языка.
2. В поле **Алфавит языка** указан алфавит языка, на основе которого Вы создаете новый язык. Если требуется, отредактируйте алфавит, нажав на кнопку **...**.
3. Словарь, который будет использоваться системой при распознавании и проверке распознанного текста. Возможны следующие варианты:
 - **Нет** (не подключать словарь к языку).
 - **Встроенный словарь** (используется словарь, поставляемый с программой).
 - **Пользовательский словарь**.

Чтобы наполнить словарь или подключить старый пользовательский словарь или текстовый файл в Windows (ANSI)-кодировке (слова должны быть разделены пробелами или другими символами, не включенными в алфавит), нажмите кнопку **Редактировать**.

Замечание: Словарные слова пользовательского языка считаются правильными, если в тексте они встретились с той капитализацией, в которой они заданы в словаре, а также в каком-либо стандартном виде: всеми маленькими, всеми большими буквами или с большой буквы:

| Вид слова в словаре: | Допустимые варианты написания слова в тексте: |
|----------------------|---|
| abc | abc, Abc, ABC |
| Abc | abc, Abc, ABC |
| ABC | abc, Abc, ABC |
| aBc | aBc, abc, Abc, ABC |

- **Регулярное выражение** (задается грамматика нового языка; подробнее о регулярных выражениях см. Регулярные выражения).

Замечания:

1. Нажав в диалоге **Свойства языков** на кнопку **Дополнительно**, Вы можете указать дополнительные свойства нового языка, например, игнорируемые внутри слова символы или исключенные из распознавания символы и т.д.
2. По умолчанию пользовательский язык будет сохранен в папку пакета. Версия ABBYY FineReader Corporate Edition позволяет указать другую папку. Более подробно о коллективной работе с пользовательскими языками и словарями к ним см. раздел “Коллективная работа с пользовательскими языками и пользовательскими словарями”.

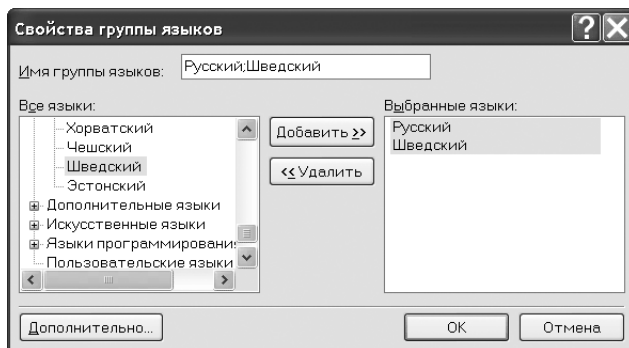
Создание новой группы языков

Если при распознавании текстов Вы часто используете некоторое сочетание языков, то Вы можете создать группу, в которую объединены эти языки. Созданная группа появится в списке языков на панели **Стандартная**.

Замечание: Вы можете указать комбинацию языков непосредственно в списке языков на панели **Стандартная**. Для этого выберите в списке строку **Выбор нескольких языков**. В открывшемся диалоге **Язык распознаваемого текста** отметьте необходимые языки.

Чтобы создать группу языков для распознавания:

1. В меню **Сервис** выберите пункт **Редактор языков...** и нажмите на кнопку **Новый**. В открывшемся диалоге выберите пункт **Создать новую группу**.
2. Перед Вами откроется диалог **Свойства группы языков**.



Что нужно определить для новой группы языков (все параметры задаются в диалоге **Свойства группы**):

1. Имя группы.
2. Подключенные языки.

Замечания:

1. Вы можете указать символы, которые заведомо не встречаются в распознаваемом документе. Указание таких символов может существенно увеличить скорость и надежность распознавания. Для этого в диалоге **Свойства группы языков** нажмите кнопку **Дополнительно** и в диалоге **Дополнительные свойства группы языков** отметьте соответствующие символы.
 2. По умолчанию пользовательская группа языков будет сохранена в папку пакета. Версия ABBYY FineReader Corporate Edition позволяет указать другую папку. Более подробно о коллективной работе с пользовательскими языками и словарями к ним см. раздел "Коллективная работа с пользовательскими языками и пользовательскими словарями".
-

Глава 7

Проверка и редактирование текста

После завершения распознавания результат появляется в окне **Текст**. Окно **Текст** – это встроенный редактор программы FineReader; в нем Вы можете проверить результаты распознавания и отредактировать распознанный текст.

Одна из возможностей текстового редактора FineReader – это встроенная проверка орфографии (список языков, для которых поддерживается проверка орфографии, приведен в разделе “Поддерживаемые языки распознавания”).

Система встроенной проверки орфографии позволяет:

1. Находить неуверенно распознанные слова (слова, в которых есть неуверенно распознанные символы).
2. Находить орфографические ошибки (неправильно написанные слова).
3. Добавлять неизвестные системе FineReader слова в словарь для того, чтобы они распознавались уверенно.

Краткое содержание раздела:

- Проверка распознанного текста
- Опции проверки и редактирования текста
- Пополнение и удаление слов из пользовательского словаря
- Редактирование текста в редакторе FineReader
- Редактирование таблиц

Проверка распознанного текста

Неуверенно распознанные символы и слова, которых нет в словаре, выделяются различными цветами. По умолчанию для выделения неуверенно распознанных символов используется голубой, для несловарных слов – розовый. Чтобы изменить цвета:

- На закладке Вид (меню **Сервис>Опции**) в поле **Объект** выберите пункт **Неуверенно распознанный символ (Несловарное слово)** и в поле **Цвет** – цвет подсветки.

Чтобы проверить результаты распознавания:

1. Нажмите кнопку 3–Проверить на панели **Scan&Read** (или выберите пункт **Проверка** в меню **Сервис**).
2. Откроется диалог Проверка.



3. В диалоге **Проверка** три окна. Верхнее окно – аналог окна “Крупный план” программы FineReader, в нем показано изображение слова с возможной ошибкой. Среднее окно показывает само слово с возможной ошибкой, в строке над этим окном выводится название типа ошибки. В нижнем окне, **Варианты**, предлагаются варианты замены данного слова (если таковые имеются). Для вариантов используется словарь, указанный в поле **Язык словаря**. Вы можете использовать любой словарь из предложенного списка.

Замечание: Для удобства проверки и редактирования текста диалог **Проверка** может быть увеличен. Для этого достаточно поместить указатель мыши на границу диалога (указатель превратится в двустороннюю стрелку). Перегасивая границу, увеличьте или уменьшите размеры диалога.



4. У Вас есть следующие возможности:

- Нажмите кнопку **Пропустить**, чтобы оставить слово, как есть.
- Нажмите кнопку **Пропустить все**, чтобы оставить все такие слова в распознанном тексте, как есть.

Замечание. При нажатии на кнопки **Пропустить** или **Пропустить все** со слова снимается “признак неуверенности” – считается, что слово не содержит нераспознанных или неуверенно распознанных символов; снимается также и визуальная подсветка таких символов в тексте. При экспорте в PDF в режиме **Заменять неуверенно распознанные слова их изображениями** слова, с которых при проверке был снят “признак неуверенности” (при нажатии кнопок **Пропустить** и **Пропустить все**), не будут заменены на свои изображения.

- Выберите вариант для замены и нажмите кнопку **Заменить** или **Заменить все**, чтобы заменить текущее слово или все такие слова в тексте. Если в окне **Варианты** нет правильного варианта для замены слова, отредактируйте слово или несколько соседних слов в среднем окне (обратите внимание, что при переходе в режим редактирования меняется назначение некоторых кнопок диалога и соответственно – надписи на них) и нажмите кнопку **Подтвердить**, чтобы зафиксировать сделанные изменения и перейти к следующему фрагменту текста.
- Нажмите **Добавить...**, чтобы добавить слово в словарь. В этом случае при дальнейшей проверке орфографии, если это слово (или одна из его форм) встретится в тексте, оно не будет считаться ошибочным.
- Нажмите **Опции...**, чтобы установить опции проверки распознанного текста.
- Нажмите **Закреть**, чтобы закрыть диалоговое окно.

Перемещение по неуверенно распознанным словам

Для более быстрой проверки результатов распознавания Вы можете воспользоваться кнопками  или  для перемещения к следующему или, соответственно, предыдущему неуверенно распознанному слову.

Также для перемещения по неуверенно распознанным словам можно использовать горячие клавиши: **F4 (Shift F4)**.

Опции проверки и редактирования текста

Устанавливаются на закладке **Проверка** (меню **Сервис>Опции**).

- Уровень выделения ошибок

Внимание! Устанавливается до запуска распознавания.

- Останавливаться на неуверенно распознанных словах
- Останавливаться на несловарных словах
- Останавливаться на сложных словах
- Пропускать слова с цифрами и другими неалфавитными символами
- Корректировать пробелы до и после знаков препинания

Уровень выделения ошибок

В списке **Уровень выделения ошибок** возможно выбрать следующие значения:

- **Нет** – ошибки распознавания не выделяются.
- **Стандартный** – цветом выделяются нераспознанные и неуверенно распознанные символы.
- **Максимальный** – помимо нераспознанных и неуверенно распознанных символов, цветом выделяются слова, которых нет в словаре языка распознавания.

Замечание: Количество показываемых в окне **Текст** ошибок будет изменено после повторного распознавания документа.

Останавливаться на неуверенно распознанных словах

Система при проверке орфографии останавливается на словах, в которых были неуверенно распознаны какие-либо буквы.

Останавливаться на несловарных словах

Эта опция позволит Вам проверить слова, которых нет в словаре и которые могли быть распознаны программой неверно.

Останавливаться на сложных словах

При проверке орфографии система останавливается на словах, которых нет в словаре, но которые могут быть построены по имеющимся морфологическим моделям или которые могут быть составлены из имеющихся в словаре слов.

Игнорировать слова с цифрами и другими неалфавитными символами

При проверке орфографии слова, внутри которых встречаются цифры или какие-либо другие символы, не входящие в алфавит языка распознавания, не считаются ошибочными, если эти слова не содержат неуверенно распознанные символы.

Корректировать пробелы до и после знаков препинания

Если этот пункт отмечен, то система не будет останавливаться на тех фрагментах, где неверно расставлены пробелы до и после знаков препинания, а исправит их автоматически.

Пополнение и удаление слов из пользовательского словаря

Добавление слова в пользовательский словарь

Добавление слов в словарь – один из способов повышения качества распознавания, т.к. при распознавании система проверяет слова по словарю. В словарь имеет смысл добавлять часто встречающиеся слова (например, термины, сокращения, названия и т.д.). При добавлении слова система строит его парадигму (совокупность всех форм данного слова). Благодаря этому система может уверенно распознавать не только ту форму, которая уже однажды встретилась в тексте и была добавлена в словарь, но и все формы добавленного слова.

Чтобы добавить слово в словарь во время проверки орфографии:

- Нажмите кнопку **Добавить** в диалоге **Проверка**.

В диалоге **Начальная форма** Вы должны установить следующие параметры:

1. **Часть речи** (Существительное, Прилагательное, Глагол, Неизменяемое слово).
2. Если слово всегда пишется с большой буквы, отметьте пункт **Имя собственное**.
Если Вы добавляете слово, являющееся сокращением, отметьте пункт **Аббревиатура**.
3. Начальная форма слова.

Нажмите **ОК**. Откроется диалог **Построение парадигмы**, в котором пользователю предлагаются вопросы, по ответам на которые строится парадигма слова. Для ответов на вопросы нажимайте кнопки **Да** или **Нет**. Если вы ошиблись при ответе на вопрос, нажмите кнопку **Сначала**, чтобы иметь возможность ответить на этот вопрос по-другому. Система покажет построенную парадигму в диалоге **Парадигма**.

Замечания:

1. Если Вы хотите, чтобы при добавлении слова в английский словарь парадигма не строилась (слово добавлялось бы как неизменяемое), то на вкладке **Проверка** (меню **Сервис>Опции**) отметьте опцию **Добавлять слово как неизменяемое**.
2. Вы можете добавлять слова во время просмотра списка добавленных слов. Для этого в меню **Сервис** выберите пункт **Просмотр словарей**. В открывшемся диалоге **Выбор словаря** выберите язык и нажмите кнопку **Просмотр**. Откроется словарь со списком добавленных слов. Вы можете, нажав на кнопку **Добавить**, добавить набранное слово.
3. Построение парадигмы при добавлении слова в словарь возможно для английского, армянского (восточный, западный и грабар), испанского, итальянского, немецкого (новая и старая орфография), русского, украинского и французского языков.

Если добавляемое слово уже есть в словаре, система предупредит Вас об этом. В этом случае Вы можете посмотреть его парадигму. Если существующая парадигма Вас не устраивает, можете создать другую (кнопка **Добавить** в диалоге **Добавить слово**).

Совет:

1. Вы можете импортировать пользовательские словари от FineReader 3.0, 4.0 или 5.0.
2. Вы можете импортировать пользовательский словарь (*.dic), который Вы создали при работе с программой Microsoft Word 6.0, 7.0, 8.0 и 9.0.

Чтобы импортировать словарь:

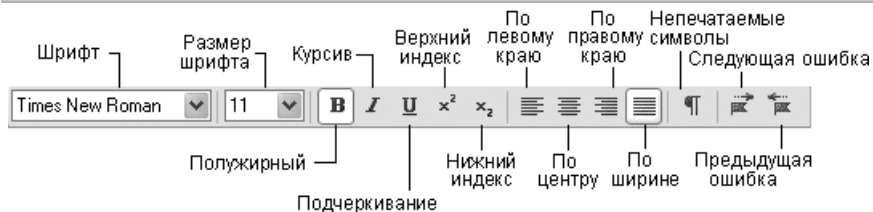
1. В меню **Сервис** выберите пункт **Просмотр словарей**. Далее выберите язык словаря и нажмите кнопку **Просмотр**.
2. В открывшемся диалоге **Словарь** нажмите кнопку **Импорт** и выберите файлы с расширениями (*.pmd, *.txt или *.dic).

Чтобы удалить слово из словаря:

1. В меню **Сервис** выберите пункт **Просмотр словарей**. Выберите язык словаря и нажмите кнопку **Просмотр**.
2. В открывшемся диалоге выделите слово и нажмите кнопку **Удалить**.

Редактирование текста в редакторе FineReader

Совет: Если символы в текстовом окне FineReader отображаются некорректно (в словах на месте некоторых букв стоят значки “?” или “□”), это означает, что шрифты, выбранные на закладке **Форматирование**, не содержат всех символов, входящих в распознаваемый язык. Установите шрифт (группа **Шрифты**, меню **Сервис>Опции**, закладка **Форматирование**), поддерживающий все символы языка документа (например, Arial Unicode MS или Bitstream Cyberbit), и перераспознайте документ. Более подробно о шрифтах для поддерживаемых языков см. раздел “Шрифты, необходимые для корректного отображения символов поддерживаемых языков в редакторе FineReader” справки программы ABBYY FineReader.





Распознанный текст выводится в окно **Текст**. При сохранении текста в приложение оформление документа сохраняется в соответствии с опциями форматирования, установленными на закладке **Форматирование** и в диалогах **Форматы**.

Неуверенно распознанные символы выделяются цветом. Вы можете отменить выделение слов с неуверенно распознанными символами: на закладке **Вид** (меню **Сервис>Опции**) снимите отметку с пункта **Выделять неуверенно распознанные символы**.

Редактор FineReader поддерживает два режима просмотра документа: режим с полным сохранением исходного оформления документа и черновой режим просмотра.

В режиме с полным сохранением исходного оформления документа блоки с распознанным текстом, таблицами и картинками отображаются именно там, где они были в исходном изображении. Таким образом, в этом режиме сохраняется исходный вид изображения: колонки, таблицы, картинки и буквицы (большие буквы на несколько строк в абзаце). Блок, в котором находится курсор, является активным. При перемещении курсора с помощью клавиатуры последовательность обхода блоков определяется их порядковыми номерами. Если текст блока не умещается внутри его границ (например, в результате редактирования распознанного текста), то часть текста неактивного блока может быть не видна на экране. Признаком этого являются специальные красные маркеры на границах блока. Когда блок становится активным, его границы расширяются таким образом, чтобы весь его текст был виден на экране целиком.

В черновом режиме редактора не отображаются: левый отступ; выравнивание абзаца (все абзацы прижаты к левому краю); цвет и фон символов. Для показа текста используется шрифт одного размера (по умолчанию 12 пт.). При этом сохраняется шрифт и форматирование текста: полужирный, курсив, подчеркнутый, верхние и нижние индексы.

Переход из одного режима в другой осуществляется нажатием кнопок  (режим отображения исходного оформления документа) и  (черновой режим редактора) в окне **Текст**.

Вы можете поменять установленный по умолчанию размер шрифта для показа в черновом режиме. Для этого:

1. В меню **Сервис** выберите пункт **Опции**.
2. На закладке **Вид** укажите нужный размер шрифта в пункте **Размер шрифта в черновом режиме**.

Редактор системы FineReader предоставляет следующие возможности по редактированию текста:

1. Копирование, перемещение, удаление выделенных фрагментов текста
2. Поиск и замена указанного фрагмента текста
3. Изменение начертания шрифта
4. Изменение выравнивания фрагмента текста
5. Отмена и восстановление действий

Копирование, перемещение, удаление выделенных фрагментов текста

1. Перед применением команд копирования, перемещения или удаления выделите нужный фрагмент текста.
2. Прделайте одно из действий:

Чтобы скопировать выделенный текст:



Кнопка
Копировать

- Нажмите кнопку **Копировать** на инструментальной панели **Стандартная**.
- В меню **Правка** или в локальном меню выберите команду **Копировать**.
- Нажмите клавиши **Ctrl+C**.

Чтобы переместить фрагмент текста:



Кнопка
Вырезать

- Нажмите кнопку **Вырезать** на инструментальной панели **Стандартная**.
- В меню **Правка** или в локальном меню выберите команду **Вырезать**.
- Нажмите клавиши **Ctrl+X**.

Чтобы вставить скопированный (вырезанный) текст:



Кнопка
Вставить

- Нажмите кнопку **Вставить** на инструментальной панели **Стандартная**.
- В меню **Правка** или в локальном меню выберите команду **Вставить**.
- Нажмите клавиши **Ctrl+V**.

Поиск и замена указанного фрагмента текста

Чтобы найти определенный фрагмент в редактируемом тексте:




1. Прделайте одну из следующих операций:
 - В меню **Правка** выберите пункт **Найти**.
 - Нажмите клавиши **Ctrl+F**.
2. В открывшемся диалоге **Поиск** в строке **Найти** укажите, что Вы хотите найти, и установите параметры поиска.

Замечание: Чтобы повторить поиск того же слова с теми же параметрами, нажмите клавишу **F3**.

Чтобы найти и заменить определенный фрагмент в редактируемом тексте:





1. Прodelайте одну из следующих операций:
 - В меню **Правка** выберите пункт **Заменить**.
 - Нажмите клавиши **Ctrl+N**.
2. В открывшемся диалоге **Заменить** в строке **Найти** укажите, что Вы хотите найти, в строке **Заменить на** укажите, на что Вы хотите заменить найденное слово, и установите параметры поиска.

Изменение начертания шрифта

1. Установите курсор на слово, которое Вы хотите изменить, или выделите участок текста, для которого Вы хотите изменить шрифт.
2. Прodelайте одну из следующих операций:
 - Нажмите на одну из кнопок , ,  на панели **Форматирование**.
 - Щелкните правой кнопкой мыши в окне **Текст** и в локальном меню выберите пункт **Шрифт**. В открывшемся диалоге **Шрифт** выберите название шрифта и установите нужные Вам параметры шрифта.
 - Нажмите клавиши **Ctrl+B** – полужирный, **Ctrl+I** – курсив, **Ctrl+U** – подчеркивание.

Замечание: Устанавливаемые в диалоге **Шрифты** цвет фона блока и форматирование текста малыми прописными (отображение всех строчных букв в выделенном тексте как прописных букв уменьшенного размера) не отображаются в окне **Текст**. Вы увидите эти изменения при сохранении документа в формате приложений, поддерживающих указанные типы форматирования текста.

Изменение выравнивания текста

1. Выделите текст, который необходимо выровнять.
2. Прodelайте одну из следующих операций:
 - Нажмите на одну из кнопок , , ,  на панели **Форматирование**.
 - Щелкните правой кнопкой мыши в окне **Текст** и в локальном меню выберите пункт **Шрифт**. В открывшемся диалоге **Шрифт** в поле **Выравнивание** выберите нужный пункт.

Отмена и восстановление действий

Прodelайте одну из следующих операций.

Чтобы отменить совершенное действие:

Кнопка

Отменить

- Нажмите кнопку **Отменить** на инструментальной панели **Стандартная**.
- В меню **Правка** выберите пункт **Отменить**.
- Нажмите клавиши **Ctrl+Z**.



Кнопка

Восстановить

Чтобы восстановить отмененное действие:

- Нажмите кнопку **Восстановить** на инструментальной панели **Стандартная**.
- В меню **Правка** выберите пункт **Восстановить**.
- Нажмите клавиши **Ctrl+Y**.

Редактирование таблиц

При редактировании таблицы Вы можете:

1. Объединить содержимое ячеек или строк
2. Разбить содержимое ячеек
3. Разбить содержимое строки (столбца)
4. Удалить содержимое ячейки

Чтобы объединить содержимое ячеек или строк:



- Удерживая клавишу **Ctrl**, выделите на изображении ячейки или строки, которые Вы хотите объединить. В меню **Правка** выберите пункт **Объединить ячейки таблицы** или **Объединить строки таблицы**.


Чтобы разбить содержимое ячеек:

- В меню **Правка** выберите пункт **Разбить ячейки таблицы**.

Внимание! Команда применяется только к ранее объединенным ячейкам таблицы.

Чтобы разбить содержимое строки или столбца:

1. На панели **Изображение** выберите инструмент  или .
2. Вставьте горизонтальную/вертикальную линию в строку/столбец таблицы, содержимое которых Вы хотите разделить.

Замечание: Воспользовавшись инструментом  или командой меню **Объединить строки таблицы** (меню **Правка**), Вы можете объединить содержимое строк (столбцов).

Чтобы удалить содержимое ячейки:

- В окне **Текст** выделите ячейку (или несколько ячеек), содержимое которой Вы хотите удалить, и нажмите клавишу **Del**.

Сохранение во внешние редакторы и форматы

Результаты распознавания можно сохранить в файл, передать во внешнее приложение, не сохраняя на диск, скопировать в буфер обмена или отправить по электронной почте в любом из поддерживаемых программой FineReader форматов сохранения. Сохранить можно все страницы или только выбранные.

Список поддерживаемых приложений:

Microsoft Word 6.0, 7.0, 97 (8.0), 2000 (9.0), 2002 (10.0) и 2003 (11.0); Microsoft Excel 6.0, 7.0, 97 (8.0), 2000 (9.0) и 2002 (10.0) и 2003 (11.0); Microsoft PowerPoint 2000 (9.0) и 2002 (10.0); Corel WordPerfect 7.0, 8.0, 9.0 и 2002 (10.0); Lotus Word Pro 9.5, 97 и Millennium Edition; StarWriter 4.x и 5.x; PROMT 98 и все приложения, поддерживающие стандарт ODMA.

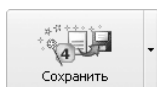
Краткое содержание главы:

- Общая информация по сохранению распознанного текста
- Опции сохранения распознанного текста
- Сохранение распознанного текста в формате RTE, DOC и Word XML
- Сохранение распознанного текста в формате PDF
- Сохранение распознанного текста в формате HTML
- Сохранение распознанного текста в формате PowerPoint
- Сохранение изображения страницы

Общая информация по сохранению распознанного текста

Вы можете:

- Сохранить распознанный текст, используя **Мастер сохранения результатов**.
- Сохранить открытую или выделенные в окне **Пакет** страницы в файл или во внешнее приложение.
- Сохранить все страницы пакета в файл или во внешнее приложение.
- Сохранить изображение страницы.



Кнопка **4–Сохранить** позволяет передать результаты распознавания в выбранное приложение или сохранить их в файл. Внешний вид значка меняется в зависимости от выбранного режима сохранения; подпись **Сохранить** меняется на название выбранного приложения.

Чтобы сохранить распознанный текст:

- Нажмите стрелку справа от кнопки **4–Сохранить** и в локальном меню выберите необходимый пункт.

Замечание: При сохранении нескольких страниц сначала выделите их в окне **Пакет**.

После того как Вы экспортировали распознанный текст в выбранное Вами приложение, отправили его по электронной почте, передали в буфер или сохранили в файл, “информация” об этом действии отразится на значке кнопки **4–Сохранить**. Поэтому для того, чтобы повторить ту же операцию для другого изображения, Вам достаточно нажать на этот значок.

Опции сохранения распознанного текста

Опции сохранения распознанного текста устанавливаются на закладке **Форматирование** диалога **Опции** (меню **Сервис>Опции**). Также часть этих опций можно установить в диалогах **Мастер сохранения результатов** и **Сохранить текст** как.

- Режим сохранения форматирования и раскладки распознанного текста
- Сохранять или не сохранять картинки в распознанном тексте
- Разрешение сохраняемых картинок
- Качество JPEG
- Используемые шрифты
- Сохранять все страницы пакета или только выделенные
- Режимы передачи страниц пакета

Режим сохранения форматирования и раскладки распознанного текста (при сохранении в форматах RTF, DOC, Word XML, PPT или HTML)

1. **Сохранять полное оформление документа** – сохраняется полное оформление документа: разбиение на абзацы, гарнитура и размер шрифта, колонки, направление текста, цвет букв и фон текста. Сохраняется структура таблиц.
2. **Сохранять начертание и размер шрифта** – сохраняется структура таблиц, разбиение на абзацы, начертание и размер шрифта.
3. **Не сохранять оформление** – в этом случае сохраняется разбиение на абзацы и структура таблиц.

Замечание: Вы можете установить дополнительные опции для каждого формата, в котором FineReader сохраняет распознанный текст. Например, для формата **RTF/DOC/Word XML** Вы можете указать формат страницы, выделять ли ошибки цветом; для формата **HTML** – разрешение картинки, кодовую страницу и т.д. Дополнительные опции сохранения устанавливаются в диалоге **Форматы** (меню **Сервис>Форматы**) на закладке, соответствующей нужному формату.

Сохранять/не сохранять картинки

Позволяет сохранить картинки в распознанном тексте. Эта опция доступна при сохранении в форматы **RTF, DOC, Word XML, PDF, PPT** или **HTML**.

Разрешение картинок (при сохранении в форматах RTF/DOC/Word XML, PDF, PPT и HTML)

Иногда может потребоваться уменьшить разрешение картинок. Например, так как HTML-файлы чаще всего просматриваются на экране, то для картинок в этих файлах нет смысла устанавливать разрешение большее, чем разрешение экрана. Таким образом, Вы можете уменьшить разрешение картинок (тем самым уменьшив размер HTML- файла) без визуальной потери качества картинок. Для этого введите требуемое значение в поле **Уменьшить разрешение картинок до** на закладке **Форматы>RTF/DOC/Word XML (PDF, PPT, HTML)**.

Замечание: Если в поле **Уменьшить разрешение картинок до** вводится значение большее, чем исходное разрешение изображения, то данное значение игнорируется программой; картинки будут сохранены с разрешением исходных изображений.

Качество JPEG (при сохранении в форматах RTF, DOC, Word XML, PDF, PPT и HTML)

При сохранении в PDF, PPT и HTML–файлы картинки сохраняются в формате JPEG. При сохранении в **RTF/DOC/Word XML** форматы Вы можете указать в качестве формата сохранения картинок JPEG формат.

В этом формате используется метод сжатия изображения с “потерей качества”. Т.е. технология сжатия построена по принципу усреднения точек группы, так что целая область записывается в виде одного числа, а не в виде большого набора цифр, описывающих каждую точку изображения в отдельности. Значение, установленное в поле **Качество JPEG**, позволяет регулировать качество сохраненного изображения (меню **Сервис>Форматы**, закладки **RTF/DOC/Word XML, PDF, PPT и HTML**). Возможно установить одно из значений в диапазоне от 1 до 100 (по умолчанию выставлено среднее значение – 50).

Чем выше значение, установленное в поле **Качество JPEG**, тем более точное изображение будет получено при сохранении. Размер сохраняемого изображения также зависит от выбранного значения: чем оно выше, тем больше размер JPEG–файла. Чтобы подобрать оптимальное сочетание качества изображения и размера файла, сохраните изображение в нескольких вариантах – с разными значениями поля **Качество JPEG** – и откройте их в какой–нибудь графической программе.

Качество JPEG устанавливается на закладке **Форматы>PDF (HTML)**.

Используемые шрифты (при сохранении в форматах RTF, DOC, Word XML, PPT или HTML)

При сохранении в файл или при передаче распознанного текста в приложение используются шрифты, установленные на закладке **Форматирование** в группе **Шрифты**. Вы можете изменить шрифты в окне **Текст** или, выбрав другие шрифты на закладке **Форматирование**, перераспознать изображение.

Сохранять все или только выделенные страницы

Позволяет сохранить все страницы пакета или только выделенные. При сохранении или экспорте части страниц сначала выделите их в окне **Пакет**.

Режимы передачи страниц пакета (при сохранении нескольких страниц пакета)

- **Каждая страница в отдельный файл** – позволяет сохранить каждую страницу пакета в отдельный файл. В этом случае к названию файла в конец автоматически добавляется порядковый номер страницы в пакете.
- **Сохранять с именами исходных изображений** – позволяет сохранить распознанные страницы в отдельные файлы выбранного формата, причем имена этих файлов будут совпадать с именами соответствующих файлов изображений.

Замечания:

1. Страницы, не связанные с исходными файлами изображений (например, полученные при сканировании), в этом режиме игнорируются, т.е. не сохраняются; при наличии в пакете таких страниц FineReader выдает перед сохранением предупредительное сообщение.
2. Если несколько страниц пакета, следующие подряд в группе выбранных сохраняемых страниц, имеют в качестве исходных изображений файлы с одинаковыми именами (или один и тот же файл), то такие страницы считаются имеющими общий источник (это могут быть различные страницы многостраничного TIFF или страницы, появившиеся в результате разрезания некоторого исходного изображения). Распознанный текст с таких страниц будет сохранен в один файл. Если же страницы с одинаковыми названиями исходных изображений в выборке следуют не подряд, такие страницы сохраняются в разные файлы, при этом к имени файла добавляется различающий индекс следующего вида: _1, _2, _3 и т.д.

- **Пофайловое деление по пустым страницам** – весь пакет рассматривается как совокупность групп страниц, разделенных пустыми страницами. Для каждой такой группы при сохранении образуется отдельный файл, в который последовательно записываются страницы группы; имена формируемых в этом режиме файлов образуются добавлением к указанному пользователем имени файла индексов следующего вида: -1, -2, -3 и т.д.
- **Все страницы в один файл** – позволяет сохранить все (или все выделенные) страницы пакета в один файл.

Сохранение распознанного текста в форматах RTF, DOC и Word XML

Внимание! Сохранение в формат Word XML возможно только при сохранении в Microsoft Word 2003.

Режимы сохранения оформления и режим сохранения картинок устанавливаются на закладке **Форматирование** диалога **Опции** (меню **Сервис>Опции**).

Замечание: При сохранении в форматах RTF, DOC и Word XML используются шрифты, установленные на закладке **Форматирование** диалога **Опции** (меню **Сервис>Опции**), или те, которые Вы выбрали в процессе редактирования текста в окне **Текст**.

Советы:

- Для более удобного редактирования распознанного текста в Microsoft Word Вы можете сохранить выделение цветом неуверенно распознанных символов. Для этого на закладке **RTF/DOC/Word XML** в группе **Выделять неуверенно распознанные символы** выберите цвет в пунктах **Цветом фона** или **Цветом символа**. В сохраненном файле ошибки будут выделены указанным Вами цветом.
 - При сохранении в формат Word XML распознанное изображение можно будет просматривать в Word в окне Крупный план. В нем будет отображаться увеличенное изображение редактируемой строки или участка документа. Пункт **Открывать окно FineReader'a Крупный план в Microsoft Word 2003** отмечен по умолчанию на закладке **RTF/DOC/Word XML**.
-

Сохранение распознанного текста в формате PDF

Опции сохранения оформления документа:

1. **Только текст и картинки** – сохраняется распознанный текст и имеющиеся в тексте картинки.
2. **Только изображение** – сохраняется только изображение.
3. **Текст поверх изображения картинки** – изображение передается картинкой. Текстовые области сохраняются как текст поверх картинки.
4. **Текст под изображением картинки** – все изображение сохраняется как картинка. Под нее “записывается” распознанный текст. В этом случае сохраняется полный дизайн документа: цветовые выделения шрифта, фона, оформление бумаги и т.д. Эта опция может использоваться, например, для составления архивов документов: полностью сохраняется исходное оформление документа, при этом появляется возможность использования функции полнотекстового поиска по ним.

Чтобы установить опции при сохранении в формате PDF:

1. В меню **Сервис** выберите пункт **Форматы**.
 2. На закладке PDF диалога **Форматы** установите требуемые опции.
-

Замечания:

1. При сохранении в режиме **Только текст и картинки** и в режиме **Текст поверх изображения страницы** Вы можете указать, чтобы при сохранении текста в формат PDF неуверенно распознанные слова заменялись их изображением. Для этого на закладке PDF диалога **Форматы** отметьте опцию **Заменять неуверенно распознанные слова их изображениями**.

2. При сохранении текстов, использующих отличную от латинской кодировку страницу (например, кириллическую, греческую, чешскую и т.д.), используются шрифты компании ПараТайп (www.paratype.com/shop).
 3. Если при сохранении текста в PDF-формате появилось сообщение о том, что текст содержит символы, отсутствующие в стандартных шрифтах, следует выбрать один из режимов для работы с True 1 шрифтами, а также сами шрифты True 1. Эти шрифты должны быть подключены через программу Adobe Type Manager или через инсталлятор postscript-шрифтов в Windows 2000. Более подробно о True 1 шрифтах см. “Использование True 1 шрифтов при сохранении в формате PDF” справки программы ABBYY FineReader.
 4. Перед редактированием в программе Adobe Acrobat PDF-файлов, использующих отличную от латинской кодировку страницу (например, кириллическую, греческую, чешскую и т.д.), следует изменить шрифт редактируемого участка текста на один из шрифтов, установленных на Вашем компьютере.
-

Сохранение распознанного текста в формате HTML

Режимы сохранения оформления устанавливаются на закладке Форматирование диалога **Опции** (меню **Сервис>Опции**).

Замечание: При сохранении в формате HTML используются шрифты, установленные на закладке Форматирование диалога **Опции** (меню **Сервис>Опции**), или те, которые Вы выбрали в процессе редактирования текста в окне **Текст**.

Чтобы сохранить картинки в HTML-файле:

- На закладке Форматирование диалога **Опции** (меню **Сервис>Опции**) отметьте опции **Сохранять картинки**.
-

Замечание: Картинки сохраняются в отдельные файлы с расширением *.jpg. Разрешение сохраняемых картинок и их качество устанавливается на закладке HTML диалога **Форматы** (меню **Сервис>Форматы**).

Форматы HTML

1. **Полный (совместим с Internet Explorer 4.0 и выше)** – файл записывается в новом формате – HTML 4, что позволяет точно передать оформление документа, используя при этом таблицу стилей, встраиваемую в HTML-файл.

2. **Простой (совместим со всеми браузерами)** – файл записывается в формате HTML 3; оформление сохраняется приблизительно (не сохраняются отступы первой строки, неточное сохранение размера шрифта). Этот формат HTML поддерживается всеми браузерами (Netscape Navigator и Internet Explorer 3.0 и выше).
3. **Авто (Простой и Полный форматы сохраняются в один файл)** – в один файл сохраняются два формата (**Простой** и **Полный**), при этом при показе файла в браузере автоматически выбирается один из них в соответствии с типом и версией браузера.

Замечание: При наличии браузеров, не полностью поддерживающих HTML4/CSS (например, Microsoft Internet Explorer 3.0 и более ранние версии, Netscape 4.x и др.), рекомендуется использовать **Простой** формат сохранения.

Чтобы установить формат HTML:

- На закладке **HTML** диалога **Форматы** (меню **Сервис>Форматы**) в группе **Форматы** установите переключатель в соответствующее положение.

Замечание: Кодовая страница определяется системой автоматически. Если Вы хотите поменять кодовую страницу, на закладке **HTML** в диалоге **Форматы** в поле **Кодовая страница** укажите нужную Вам страницу.

Сохранение распознанного текста в формате PowerPoint

Режимы сохранения оформления устанавливаются на закладке **Форматирование** диалога **Опции** (меню **Сервис>Опции**).

Замечание: При сохранении в формате PPT используются шрифты, установленные на закладке **Форматирование** диалога **Опции** (меню **Сервис>Опции**), или те, которые Вы выбрали в процессе редактирования текста в окне **Текст**.

Внимание! При сохранении в формат PPT результаты распознавания сохраняются в виде специальных HTML-файлов, содержащих разные части презентации. Для того чтобы презентация была сохранена в одном файле, ее необходимо повторно сохранить в программе PowerPoint, выбрав в меню **Файл** пункт **Сохранить как** и указав в качестве формата сохранения формат PPT.

Сохранение изображения страницы

1. Выделите страницу в пакете.
2. В меню **Файл** выберите пункт **Сохранить изображение как**.
3. В открывшемся диалоге **Сохранить изображение как** выберите диск, папку для размещения сохраняемого файла и формат.
Замечание. Вы можете ограничить сохраняемую в файле область изображения рамками выделенных на нем блоков (любого типа). Для этого выделите нужный блок или блоки, а затем в диалоге **Сохранить изображение как** отметьте пункт **Сохранять выделенные блоки**. Этот режим работает только при сохранении одиночного изображения.
4. Дайте имя сохраняемому файлу.
5. Нажмите **Save**.

Совет: Вы можете сохранить несколько изображений в один файл как многостраничный TIF. Для этого:

1. Выделите нужные изображения в окне **Пакет**.
2. В меню **Файл** выберите пункт **Сохранить изображение как**. В открывшемся диалоге выберите формат TIF и отметьте опцию **Сохранять все страницы в один файл**.

Замечание: Если Вы сохраняете несколько выделенных в окне **Пакет** изображений в виде отдельных файлов (т.е. не используя режим сохранения в многостраничный TIF), имена этих файлов будут образовываться добавлением к введенному Вами имени файла номера страницы в виде четырехзначного числа.

Коллективная работа в сети

ABBYY FineReader Corporate Edition имеет возможности для коллективной работы в сети. На каждом компьютере должна быть установлена отдельная копия программы (более подробную информацию по установке программы FineReader на сервер и на рабочие станции Вы можете найти в “Руководстве системного администратора”, которое находится в подпапке **Administrator’s Guide** папки сервера, куда был установлен FineReader). Эта версия имеет следующие возможности для работы на нескольких компьютерах:

1. Работа с одним и тем же пакетом на нескольких компьютерах.

В этом режиме повышается производительность массовой обработки документов; имеется возможность отслеживания процесса обработки страниц – кем (имя пользователя и имя компьютера) в данный момент страница открыта, кем она была отсканирована, распознана, проверена и т.д. Изменения, производимые со страницей одним пользователем, видны всем пользователям, работающим с этим пакетом.

2. Коллективная работа с пользовательскими языками и пользовательскими словарями

ABBYY FineReader Corporate Edition позволяет использовать одни и те же пользовательские языки и подключенные к ним пользовательские словари нескольким пользователям одновременно. При этом пользовательские словари могут пополнять несколько пользователей, например, в процессе проверки результатов распознавания.

3. Коллективная работа со вспомогательными словарями к языкам со словарной поддержкой

Для языков со словарной поддержкой в ABBYY FineReader имеются встроенные словари. Эти словари содержат наиболее часто используемые слова, однако в них могут отсутствовать имена собственные, узкоспециальные термины, сокращения и т.п. Для повышения качества распознавания (при распознавании система проверяет слова по словарю) и для ускорения процедуры проверки распознанного текста Вы можете добавлять такие слова во вспомогательный словарь. Программа ABBYY FineReader Corporate Edition позволяет нескольким пользователям работать с одним и тем же вспомогательным словарем, в том числе и пополнять его.

Краткое содержание раздела:

- Коллективная работа с одним пакетом на нескольких компьютерах
- Коллективная работа с пользовательскими языками и пользовательскими словарями
- Коллективная работа со вспомогательными словарями к языкам со словарной поддержкой

Коллективная работа с одним пакетом на нескольких компьютерах


(только для версии ABBYY FineReader Corporate Edition)

1. Создайте/Откройте пакет и установите для созданного пакета нужные опции сканирования и распознавания.
Запустите FineReader и откройте пакет на всех машинах, которые будут с ним работать.
2. На машинах, предназначенных для распознавания, запустите Распознавание в фоновом режиме (меню **Процесс>Запустить фоновое распознавание**).
3. Запустите сканирование на машине, где установлен сканер с автоподатчиком страниц.

Совет: Если Ваш скоростной сканер не поддерживает работу с TWAIN-драйвером, то Вы можете отсканировать документы прямо в папку пакета программы FineReader. Для этого на машине, к которой подключен сканер с автоподатчиком, отсканируйте изображения любой графической программой. При этом для хранения отсканированных изображений укажите именно ту папку, в которой хранится пакет, созданный в программе FineReader. Отсканированные файлы должны именоваться специальным образом (в порядке сканирования): 0001.tif, 0002.tif, 0003.tif...
FineReader автоматически “подхватит” и обработает все отсканированные изображения.

4. Распознанный текст Вы можете отредактировать и сохранить в файл или в выбранное приложение.

В окне **Пакет** Вы можете отслеживать состояние страниц. Например, кем (имя пользователя и имя компьютера) данная страница в данный момент открыта; была ли страница распознана, отредактирована и т.д. Вся эта информация отображается в соответствующих колонках окна **Пакет** в режиме отображения страниц **Вид со свойствами**. Чтобы установить этот режим отображения страниц:

- Нажмите кнопку  на панели **Стандартная** или
- Щелкните правой кнопкой мыши в окне **Пакет** и в локальном меню выберите пункт **Вид>Вид со свойствами**.

Вы можете изменить настройки режима **Вид со свойствами**. Например, выбрать отображаемые в окне **Пакет** колонки или выбрать колонку, по которой будут сортироваться страницы. Чтобы изменить настройки:

- Щелкните правой кнопкой мыши в окне **Пакет** и в локальном меню выберите пункт **Вид>Настройка**. В открывшемся диалоге **Настройки окна Пакет** выберите закладку **Вид со свойствами**.

При обработке страниц на нескольких компьютерах FineReader распределяет работу между ними автоматически: появившуюся отсканированную страницу “подхватывает” первый освободившийся компьютер, на котором запущено Фоновое распознавание. При этом для других компьютеров данная страница блокируется. Обновить список страниц пакета можно, нажав **F5** или выбрав в меню **Пакет** пункт **Обновить список страниц**. По мере распознавания страниц Вы можете открывать их для проверки, редактирования и сохранения. Изменения, производимые со страницей одним пользователем, видны всем пользователям, работающим с этим пакетом.

Замечание: На многопроцессорных компьютерах режим **Фоновое распознавание** позволяет увеличить скорость распознавания пакетов, содержащих большое количество страниц.

Коллективная работа с пользовательскими языками и пользовательскими словарями

(только для версии ABBYY FineReader Corporate Edition)

Создайте пакет и установите для созданного пакета нужные опции сканирования и распознавания.

Пользовательские словари хранятся в той же папке, что и пользовательские языки. По умолчанию это папка пакета. Перед созданием пользовательского языка укажите папку хранения пользовательских языков с их словарями. Это может быть как папка на локальном диске, так и сетевая папка. Чтобы указать папку хранения пользовательских языков:

- Нажмите кнопку **Изменить** в диалоге **Редактор языков** (меню **Сервис>Редактор языков**) и в открывшемся диалоге выберите требуемую папку.

Все создаваемые пользовательские языки и пользовательские словари, подключенные к этим языкам, будут сохраняться программой в этой папке.

После того как настройка пакета завершена, сохраните настройки пакета в файл шаблона пакета (*.fbt). Для этого:

- Нажмите кнопку **Сохранить** на закладке **Опции>Распознавание** (меню **Сервис>Опции**). В открывшемся диалоге **Сохранить шаблон пакета как** выберите диск, папку для размещения сохраняемого файла и введите имя файла.

Чтобы несколько пользователей могли работать с созданными пользовательскими языками и подключенными к ним пользовательскими словарями, каждый пользователь, создавая новый пакет, должен загрузить настройки пакета из ранее сохраненного FBT-файла.

Для этого в поле **Шаблон пакета** выберите пункт **Шаблон пакета (.fnt)**. В открывшемся диалоге **Открыть шаблон пакета** выберите нужный FNT-файл. В этом случае для нового пакета будут установлены опции, ранее сохраненные в FNT-файл. Таким образом, для всех пользователей будет установлен общий путь к папке хранения пользовательских языков с их словарями.

Применяя эти языки при распознавании, Вы можете пополнять и редактировать их словари. Изменения, сделанные одним пользователем, становятся доступными и всем остальным, работающим с этой папкой. Также пользовательские языки, созданные в этой папке, доступны всем пользователям, применяющим при создании пакета этот же шаблон. Список доступных пользовательских языков и их параметры Вы можете посмотреть в диалоге **Редактор языков** в группе **Пользовательские языки**.

При добавлении, изменении или удалении слова словарь блокируется пользователем. Обновление словаря происходит при нажатии кнопки **Добавить** в диалоге **Проверка результатов распознавания** или любой кнопки в диалоге **Просмотр словарей**.

Замечания:

1. При использовании несколькими пользователями папки, в которой хранятся пользовательские языки и словари, каждый пользователь должен иметь доступ на чтение и запись в эту папку.
2. При совместном использовании пользовательские языки доступны в режиме “чтения”. Т.е. Вы не сможете изменить параметры уже существующего пользовательского языка. При этом Вы сможете пополнять и удалять слова из его пользовательского словаря.

Коллективная работа со вспомогательными словарями к языкам со словарной поддержкой

(только для версии ABBYY FineReader Corporate Edition)

Создайте пакет и установите для созданного пакета нужные опции сканирования и распознавания. По умолчанию для хранения вспомогательных словарей для языков со словарной поддержкой используется папка, в которую была установлена программа (в Windows 2000 – Documents and Settings\[user profile]\Application Data\ABBYY\FineReader\7.00\UserDictionaries).

Чтобы несколько пользователей могли работать с одним комплектом вспомогательных словарей при распознавании и проверке, необходимо установить общую папку в качестве папки хранения вспомогательных словарей. Это может быть папка на Вашем диске или сетевая папка. Для этого:

- На закладке **Проверка** диалога **Опции** (меню **Сервис>Опции**) нажмите кнопку **Изменить**. В открывшемся диалоге укажите требуемую папку.

Каждый из пользователей может пополнять эти вспомогательные словари. На время добавления, изменения или удаления слова словарь блокируется пользователем. Изменения, производимые во вспомогательном словаре одним пользователем, становятся доступными другим пользователям после обновления словаря. Обновление словаря происходит при нажатии кнопки **Добавить** в диалоге **Проверка результатов распознавания** или любой кнопки в диалоге **Просмотр словарей**.

Замечание: При использовании несколькими пользователями папки, в которой хранятся вспомогательные словари, каждый пользователь должен иметь доступ на чтение и запись в эту папку.

Работа с пакетами

В пакете хранятся как исходные изображения, так и соответствующий им распознанный текст. Большинство установок FineReader сохраняются на пакет (опции сканирования, распознавания, сохранения, а также созданные в процессе работы пользовательские эталоны, языки и группы языков). При создании нового пакета могут использоваться настройки по умолчанию, настройки текущего пакета или настройки из файла с расширением *.fbt.

Краткое содержание раздела:

- Общая информация по работе с пакетом
- Создание нового пакета
- Открытие пакета
- Добавление изображений в пакет
- Нумерация страниц пакета
- Сохранение пакета
- Закрывание страницы пакета или всего пакета
- Удаление пакета
- Настройки пакета
- Поиск текста по распознанным страницам пакета

Общая информация по работе с пакетом



При запуске программы FineReader по умолчанию открывается новый пакет. Вы можете работать с этим пакетом или создать новый. В пакете может содержаться до 9999 страниц. В один пакет для удобства работы рекомендуется объединять изображения, логически связанные между собой (например, страницы одной книги, тексты на одном языке или изображения с однотипным расположением текста и т.д.).

В окне **Пакет** Вы можете видеть список страниц, входящих в открытый пакет. Чтобы просмотреть страницу, нужно щелкнуть мышью по ее значку или дважды щелкнуть по ее номеру. Откроются файлы, которыми данная страница представлена в пакете: текстовый (если данная страница была распознана) и графический, – в окнах **Текст** и **Изображение** соответственно.

Существует два основных способа представления страниц в окне **Пакет**:

| Способ представления | Описание |
|--------------------------|---|
| Пиктограммы | Страницы пакета отображаются пиктограммами – своими сильно уменьшенными изображениями. По мере обработки изображения на пиктограмме появляются дополнительные специальные значки, отображающие действия, произведенные над страницей. Этот способ представления страниц пакета удобно использовать, например, для открытия необходимой страницы пакета: страница представлена своим миниатюрным изображением, и Вам не надо запоминать номер, под которым она была отсканирована. Чтобы открыть изображение, щелкните мышью по его пиктограмме. |
| Вид со свойствами | В этом случае в окне пакета отражается дополнительная информация о странице и возможна сортировка по выбранному признаку. Вид со свойствами удобен при обработке пакета, содержащего большое количество страниц, т.к. на экране помещается большее (чем в режиме Пиктограммы) количество страниц. Чтобы открыть изображение, дважды щелкните на соответствующий ему значок в окне Пакет . |

Чтобы выбрать вид страниц в окне **Пакет**:

- Нажмите на соответствующую кнопку ( / ) на панели инструментов **Стандартная** или
- Щелкните правой кнопкой мыши в окне **Пакет** и в открывшемся локальном меню выберите пункт **Вид>...**

Помимо выбора способа представления изображений в окне **Пакет**, Вы можете их настроить: выбрать показываемые свойства страницы, способ сортировки страниц пакета и т.д. Для этого:


- Щелкните правой кнопкой мыши в окне **Пакет** и в открывшемся локальном меню выберите пункт **Вид>Настройка**. В открывшемся диалоге на вкладках Пиктограммы и Вид со свойствами установите нужные Вам опции.

В одном пакете можно выделить несколько страниц подряд или выборочно.

- **Чтобы выделить несколько страниц подряд:** удерживая клавишу **Shift**, укажите с помощью мыши первую, а затем последнюю страницу из выборки.
- **Чтобы выделить несколько страниц выборочно:** удерживая клавишу **Ctrl**, укажите на них мышью.
- **Чтобы выделить все страницы пакета:** при активном окне пакета из меню **Правка** выберите пункт **Выделить все** или нажмите **Ctrl+A**.

Создание нового пакета

Чтобы создать новый пакет:


1. В меню **Файл** выберите пункт **Новый пакет** или
2. Нажмите кнопку **Новый пакет** ().

Открытие пакета


При запуске FineReader по умолчанию открывает новый пакет.

Замечание: Если Вы хотите, чтобы при запуске программа открывала последний пакет, с которым Вы работали, отметьте опцию **Открывать последний пакет при запуске FineReader** на закладке **Общие** диалога **Опции** (меню **Сервис>Опции**).

Чтобы открыть другой пакет:

1. В меню **Файл** выберите пункт **Открыть пакет** или нажмите кнопку **Открыть пакет** ().
2. В открывшемся диалоге **Открыть пакет** выберите папку, где находится нужный Вам пакет.
В случае, если Вы работали с новым пакетом, и он Вам понадобится для дальнейшей работы, сохраните его предварительно перед выходом из программы.

Вы можете открыть пакет из Windows Explorer. Для этого:

- Щелкните правой кнопкой мыши по папке пакета (отображается значком ) и в локальном меню выберите пункт **Открыть с помощью FineReader**. Запустится новая копия FineReader с выбранным пакетом.

Добавление изображений в пакет

- В меню **Файл** выберите пункт **Открыть изображение**.
Клавиатурная команда: **CTRL+O**.
- В диалоге **Открыть изображение** выберите одно или несколько изображений.
Изображение будет добавлено в открытый пакет, и его копия будет сохранена в папке пакета.

Замечание: Вы можете добавить изображение из Windows Explorer. Для этого:

1. В Windows Explorer выберите один или несколько файлов с изображениями.
2. Щелкните правой кнопкой мыши и в локальном меню выберите пункт **Открыть с помощью FineReader**. Если на Вашем компьютере уже открыт FineReader, выбранные файлы будут добавлены в текущий пакет, в противном случае автоматически запустится копия FineReader и файлы будут добавлены в новый пакет.
Данный пункт локального меню доступен только для файлов, формат которых поддерживается FineReader 7.0.

Нумерация страниц пакета

Все страницы в пакете пронумерованы. Один пакет может содержать до 9999 страниц (четырёхзначный номер). Номер страницы указывается непосредственно в самом пакете, около условного обозначения страницы.

Страницы можно перенумеровать непосредственно в окне **Пакет** или через диалог **Перенумеровать страницы**.

Чтобы перенумеровать страницы в окне **Пакет:**

1. Выделите страницу в окне **Пакет** и щелкните мышью по ее номеру или нажмите F2.
2. Введите новый номер страницы.

После изменения номера положение страницы в окне **Пакет** изменится в соответствии с введенным номером.

Замечани: Двойной щелчок мыши по номеру страницы открывает эту страницу в программе.

Чтобы перенумеровать страницы через диалог **Перенумеровать страницы:**

1. Выделите страницу или несколько страниц.
 2. В меню **Пакет** выберите пункт **Перенумеровать страницы...**
 3. Введите новый номер для первой страницы из выборки (страницы с наименьшим номером).
-

Замечания:

1. Если Вы хотите перенумеровать все страницы пакета, то в диалоге **Перенумеровать страницы** выберите пункт **Все страницы**.
2. Если Вы хотите перенумеровать часть страниц пакета, то
 - В окне **Пакет** выделите те страницы, которые Вы хотите перенумеровать.
 - В диалоге **Перенумеровать страницы** выберите пункт **Только выделенные**.
3. Если Вы хотите, чтобы выделенные страницы были перенумерованы по порядку, начиная с указанного номера, то отметьте опцию **Сплошная нумерация страниц**. Например, перенумеровываются страницы 2, 5, 6; в качестве начального номера был указан 1. Новые номера страниц будут: 1, 2, 3. В противном случае (опция **Сплошная нумерация страниц** не отмечена) новыми номерами страниц будут: 1, 4, 5. Т.е. первая из участвующих в операции страниц получит указанный Вами начальный номер, а остальные будут перенумерованы таким образом, чтобы сохранить интервалы нумерации между страницами.
4. Если перенумерация затрагивает не все страницы пакета, то возможна ситуация, когда в соответствии с заданными Вами параметрами некоторая страница должна получить новый номер, совпадающий с номером одной из страниц, не участвующих в операции. В этом случае FineReader выдаст соответствующее предупреждение, и вся операция будет отменена.

Сохранение пакета

Чтобы сохранить пакет:

- В меню **Файл** выберите пункт **Сохранить пакет как**.
- В открывшемся диалоге **Сохранить пакет как** укажите название пакета и путь, по которому будет храниться данный пакет.

Заккрытие страницы пакета или всего пакета

Чтобы закрыть страницу пакета:

- В меню **Пакет** выберите пункт **Закреть текущую страницу**.

Чтобы закрыть пакет:

- В меню **Файл** выберите пункт **Закреть пакет**.

Удаление пакета

Внимание! При удалении пакета удаляются все его страницы (изображения и текст) и дополнительные файлы, созданные в процессе работы с этим пакетом: эталон, пользовательские языки. Т.е. удаляется все содержимое папки, соответствующей пакету.

- Чтобы удалить пакет, в меню **Пакет** выберите пункт **Удалить пакет**.

Чтобы удалить страницу в пакете:

1. Выделите нужные страницы в окне пакета.
2. В меню **Пакет** выберите пункт **Удалить страницу**.
Клавиатурная команда: **DEL**.

Настройки пакета

Чтобы сохранить настройки пакета в отдельный файл:

- На закладке **Общие** (меню **Сервис>Опции**) нажмите кнопку **Сохранить**.
- В открывшемся диалоге укажите имя файла.

В этот файл будут сохранены опции с закладок Сканирование/Открытие, Форматирование, Распознавание и Проверка; опции с закладок диалога **Форматы**; путь к папке, в которой хранятся пользовательские языки и их словари; а также пользовательские языки, группы языков, пользовательские словари и эталоны. Если Вы хотите, чтобы данный

шаблон применялся ко всем новым пакетам, отметьте опцию **Применять данный шаблон к новым пакетам** в диалоге сохранения шаблона пакета.

Чтобы вернуться к опциям, устанавливаемым системой по умолчанию:

- На закладке **Общие** нажмите кнопку **Вернуть**.

Чтобы загрузить опции:

- На закладке **Общие** нажмите кнопку **Загрузить** и выберите **Шаблон пакета FineReader (*.fnt)**, содержащий требуемые опции.

Поиск текста по распознанным страницам пакета

Внимание! Для работы данной опции необходим Internet Explorer 4.0 или выше.

Вы можете задать поиск текстового фрагмента по всем страницам пакета, содержащим распознанный текст. В качестве искомого текстового фрагмента может выступать одно слово или несколько слов. При этом слово или слова могут быть в любой форме (для языков со словарной поддержкой); слова из запроса могут находиться в тексте страницы на произвольном расстоянии друг от друга и располагаться в произвольном порядке.

Для этого:

1. В меню **Правка** выберите пункт **Расширенный поиск** или нажмите **Alt+F3**.
2. Откроется окно **Поиск** (под окном **Крупный план**).
3. В поле **Найти** введите, вставьте из буфера обмена или выберите из списка ранее введенных Вами запросов искомый текст.
4. Нажмите кнопку **Найти**.

В окне **Поиск** появится список номеров страниц пакета, на которых были найдены ВСЕ искомые слова. Дополнительно для каждой найденной страницы указывается дата ее последней модификации и ближайший к началу страницы фрагмент ее текста, включающий одно или несколько из искомых слов (выделены цветом). Щелчок по номеру страницы открывает ее в окнах **Изображение**, **Текст** и **Крупный план**; при этом все найденные слова из запроса выделяются цветом фона во всех трех окнах.

Замечание: Расширенный поиск не позволяет осуществлять поиск специальных символов (например, символа конца строки или конца абзаца).

Приложение

Горячие клавиши и глоссарий

Краткое содержание раздела:

- Горячие клавиши
- Глоссарий

Горячие клавиши

| Меню | Команда | Горячие клавиши |
|--------|--|----------------------|
| Файл | Открыть изображение | Ctrl+O |
| | Сканировать изображение | Ctrl+K |
| | Сканировать несколько страниц | Ctrl+Shift+K |
| | Остановить сканирование | Ctrl+T |
| | Создать новый макет | Ctrl+N |
| | Открыть пакет | Ctrl+P |
| | Сохранить текст в файл | Ctrl+F2 |
| | Сохранить изображение в файл | F12 |
| Правка | Отменить последнее действие | Ctrl+Z |
| | Восстановить последнее отмененное действие | Ctrl+Y |
| | Удалить выделенный фрагмент и поместить его в буфер обмена | Ctrl+X |
| | Скопировать выделенный фрагмент в буфер обмена | Ctrl+INS или Ctrl+C |
| | Вставить содержимое буфера обмена | Ctrl+V или Shift+INS |
| | Удалить активный блок, фрагмент текста или выделенный в окне Пакет страницы | DEL |
| | Выделить весь текст в окне Текст, все страницы в окне Пакет или все блоки на открытой странице | Ctrl+A |
| | Найти указанный текст | Ctrl+F |
| | Найти следующее вхождение указанного текста | F3 |
| | Найти и заменить указанный текст | Ctrl+H |
| Вид | Увеличить изображение в окне Изображение | Ctrl+Shift+NUM+ |
| | Уменьшить изображение в окне Изображение | Ctrl+Shift+Num – |
| | Показать активный блок | Ctrl+Shift+Num * |
| | Свойства | Alt+ENTER |
| Пакет | Открыть следующую страницу пакета | Alt+Down |
| | Открыть предыдущую страницу пакета | Alt+Up |
| | Открыть страницу с указанным номером | Ctrl+G |
| | Закрыть текущую страницу | Ctrl+4 |
| | Удалить распознанный текст в окне Текст | Ctrl+Shift+DEL |
| | Удалить все блоки в окне Изображение и весь распознанный текст в окне Текст | Ctrl+DEL |
| | Обновить список страниц | F5 |

| Меню | Команда | Горячие клавиши |
|----------------------------------|--|------------------------|
| Процесс | Сканировать и распознать изображение | Ctrl+D |
| | Открыть и распознать изображение | Ctrl+Shift+D |
| | Сканировать и распознать изображение в режиме Мастер Scan&Read | Ctrl+W |
| | Анализировать макет страницы | Ctrl+E |
| | Анализировать все страницы пакета | Ctrl+Shift+E |
| | Распознать активную или выделенные страницы пакета | Ctrl+R |
| | Распознать все страницы пакета | Ctrl+Shift+R |
| | Распознать активный или выделенные блоки | Ctrl+Shift+B |
| Сервис | Проверить распознанный текст | F7 |
| | Перейти к следующей ошибке или неуверенно распознанному символу | F4 |
| | Перейти к следующей ошибке или неуверенно распознанному символу | Shift+F4 |
| | Просмотреть и отредактировать словарь | Ctrl+Shift+V |
| | Перевести слово в Lingvo | Ctrl+Shift+T |
| | Открыть диалог Редактор языков, в котором Вы можете создать и отредактировать язык или группу языков | Ctrl+Shift+L |
| | Открыть диалог Редактор эталонов, в котором Вы можете создать и отредактировать пользовательские эталоны | Ctrl+Shift+A |
| | Установить параметры сканирования | Ctrl+Shift+S |
| | Открыть диалог Форматы, в котором Вы можете установить опции сохранения для различных форматов | Ctrl+Shift+X |
| | Открыть диалог Опции | Ctrl+Shift+O |
| | Окна | Открыть следующее окно |
| Открыть предыдущее окно | | Ctrl+Shift+F6 |
| Открыть окно Пакет | | Alt+1 |
| Открыть окно Изображение | | Alt+2 |
| Открыть окно Текст | | Alt+3 |
| Открыть окно Крупный план | | Alt+4 |
| Перейти в окно Расширенный поиск | | Alt+5 |
| Открыть окно Расширенный поиск | | Alt+F3 |
| Справка | Открыть справку | F1 |

| Меню | Команда | Горячие клавиши |
|-------------|--|-------------------------------------|
| Общие | Отметить выделенный фрагмент текста как полужирный | Ctrl+B |
| | Отметить выделенный фрагмент текста как курсив | Ctrl+I |
| | Подчеркнуть выделенный фрагмент текста | Ctrl+U |
| | Перейти к следующей ячейке таблицы | Стрелки влево, вправо, вниз и вверх |

Глоссарий

А

ADF (Automatic Document Feeder)

устройство автоматической подачи бумаги, позволяющее отсканировать большое количество документов без ручного вмешательства. FineReader поддерживает сканирование многостраничных документов.

dpi (Dots per Inch)

количество точек на дюйм; единица измерения разрешения.

Hot folder

специальный режим открытия отсканированных изображений, при котором эти изображения будут автоматически открываться в окне программы по мере их поступления в указанную пользователем папку.

Open&Read

запускает обработку изображения: открытие, анализ макета страницы и распознавание.

Scan&Read

главная кнопка программы, по нажатию на которую система сканирует изображение и распознает (анализирует макет страницы и распознает документ).

TWAIN, TWAIN-диалог – диалог сканера.

Unicode

стандарт, разработанный концерном Unicode. Управляет процессом кодирования символов и предоставляет шестнадцатититную международную расширяемую систему кодировки символов для обработки информации на большинстве языков народов мира. Стандарт Unicode определяет кодировку символа, а также свойства и алгоритмы, используемые при реализации этого процесса.

А

Аббревиатура

сокращение, образованное из начальных букв слов. Например, МГУ, MS-DOS и т.д.

Активация

процесс получения пользователем в компании АBBYY специального кода, необходимого для обеспечения работы программы в полнофункциональном режиме на конкретном компьютере.

Активационный код

код, который выдается пользователю после прохождения процедуры активации (для версии Professional Edition). Предназначен для активации продукта на той машине, на которой был получен Installation ID.

Активационный файл

файл, который выдается пользователю после прохождения процедуры активации (для версии Corporate Edition). В нем содержится информация, необходимая для активации сервера или компьютера (в случае установки на отдельное рабочее место). Рабочие станции активируются сервером.

Автоподбор яркости

автоматический подбор яркости, производится сканером или системой FineReader. Автоподбор позволяет подбирать яркость для каждого участка изображения в отдельности.

Активный блок

блок, над которым выполняется большинство производимых действий (например, удаление, изменение типа и т.д.). На углах такого блока имеются “квадратики”.

Анализ макета страницы

(выделение блоков)

процесс выделения блоков на изображении. Блоки могут быть разных типов. Анализ макета может проводиться автоматически одновременно с распознаванием при нажатии кнопки **2-Распознать** или вручную до распознавания.

Б

Блок

участок изображения, выделенный в рамку.

Блок Картинка

этот блок используется для обозначения картинок. Он может содержать картинку или любую другую часть текста, которую Вы хотите передать в распознанный текст в качестве картинки.

Блок Штрих-код

(только в версии Corporate Edition)

этот блок используется для распознавания штрих-кодов. Т.е., если Ваш документ содержит штрих-код и Вы хотите передать его не в качестве картинки, а перевести его в последовательность букв и цифр, то выделите штрих-код в отдельный блок и присвойте ему тип **Штрих-код**.

Замечание: По умолчанию опция, позволяющая искать и распознавать штрих-коды, отключена. Чтобы подключить ее, отметьте пункт **Выделять штрих-коды** на закладке **Расознавание** (меню **Сервис>Опции**).

Д

Драйвер

программа, управляющая устройством.

З

Запрещенные символы

в качестве запрещенных символов для текущего языка указываются те, которые заведомо не могут встречаться в текстах,

распознаваемых с подключением данного языка. Указание таких символов может существенно увеличить скорость и надежность распознавания. Например, при распознавании текстов, в которых встречаются только заглавные буквы, в качестве запрещенных следует указать все строчные буквы.

Зона распознавания

блок используется для распознавания и автоматического анализа части изображения. После нажатия на кнопку **2-Распознать** выделенный блок автоматически анализируется и распознается.

И

Игнорируемые символы

в качестве игнорируемых символов указываются те, которые могут встречаться внутри слова, например, знаки слогаделения или ударения в словарях. При проверке по словарю программа не учитывает эти знаки. В распознанном тексте эти символы сохраняются, но при проверке орфографии не учитываются.

Инvertированное изображение

изображение, текст которого напечатан на темном фоне светлыми буквами.

К

Кодовая страница

таблица, в которой задано отношение между кодами символов и их начертаниями. В системе FineReader кодовая страница представлена как набор символов, из которого можно выбрать нужные символы.

Л

Лигатура

пара или тройка букв, которые из-за особенностей начертания склеиваются между собой. Например, ge, go, fe, ffi, tt и т.п.

М

Мастер Scan&Read

специальный режим сканирования и распознавания, в котором система контролирует действия пользователя и подсказывает ему, что делать, чтобы получить тот или иной результат.

Макет страницы

расположение текста, таблиц и картинок в документе, разбиение на абзацы, гарнитура и размер шрифта, колонки, направленные текста, цвет букв и фон текста.

Менеджер лицензий

приложение, в котором осуществляется управление лицензиями и активация ABBYY FineReader 7.0 Corporate Edition.

Моноширинный шрифт

любой шрифт; буквы которого имеют одинаковую ширину (моно). Например, шрифт Courier New. Для улучшения качества распознавания моноширинных шрифтов на закладке Распознавание в пункте **Тип печати** установите переключатель в положение **Пишущая машинка**.

Мягкий перенос

перенос (¬) показывает, в каком именно месте должно быть разорвано слово или словосочетание (например, “Автоформат”), если оно попадает на конец строки (например, “Авто–формат”). Все переносы в словарных словах FineReader заменяет на мягкий перенос.

Н

Начертание шрифта

способ выделения в тексте (полужирный, наклонный, с подчеркиванием, перечернутый, верхний индекс, нижний индекс, малые прописные).

Начальная форма

форма, в которой слова даются в словаре. Для существительных это форма именительного падежа единственного числа; для прилагательных – форма мужского рода единственного числа именительного падежа; для глаголов, причастий и деепричастий – это неопределенная форма (инфинитив).

О

Обучение

создание пар “растровое изображение – название символа”.

Омнифонтовая система

система распознавания, которая распознает символы практически любых размеров и начертаний.

Очистка изображения

удаление отдельных мелких черных точек на изображении.

П

Пакет

это папка, в которой хранятся изображения и рабочие файлы программы. В пакете может содержаться до 9999 страниц. В один пакет для удобства работы рекомендуется объединять изображения, логически связанные между собой (например, страницы одной книги, тексты на одном языке или изображения с однотипным расположением текста и т.д.).

Парадигма

совокупность всех грамматических форм слова.

Префиксы

небуквенные символы, которые могут встречаться в абсолютном начале слова.

Принцип одной кнопки

принцип, когда при нажатии на одну кнопку (Scan&Read) система сканирует и распознает документ.

Р

Разделители

символы, которые могут разделять слова, например, /, \, тире и т.п. и которые пишутся отдельно от слов.

Разрешение

параметр сканирования, показывает кол-во точек на единицу длины. 300 dpi (размер шрифта 10 и более пунктов), 400–600 dpi для текстов, набранных мелким шрифтом (9 и менее пунктов).

С

Сканер

устройство, предназначенное для ввода изображений в компьютер.

Сложное слово

слово, отсутствующее в словаре, но которое может быть образовано из имеющихся в словаре слов.

Суффиксы

небуквенные символы, которые могут встречаться в абсолютном конце слова.

Т

Табличный блок

этот блок используется для обозначения таблиц или текста, имеющего табличную структуру. При распознавании программа разбивает данный блок на строки и столбцы и формирует табличную структуру. В выходном тексте данный блок передается таблицей. Вы можете выделить и отредактировать таблицу вручную.

Текстовый блок

блок используется для обозначения текста. Он должен содержать только одноколоночный текст. Если внутри текста содержатся картинки, выделите их в отдельные блоки.

Тип изображения

параметр сканирования; изображение может быть черно-белым, серым или цветным.

Тип (печати) входного текста

особенность в начертании символов входного текста, в зависимости от того, каким образом он напечатан (в типографии, на матричном принтере в черновом режиме, на пишущей машинке). Для типографского текста следует устанавливать режим

Авто, для машинописного – режим **Пишущая машинка**, для текста, напечатанного на матричном принтере в черновом режиме – режим **матричный принтер**.

Тип блока

блок может быть Зоной распознавания, Текстовым, Картинкой, Таблицей, Штрих-кодом.

Ф

Фоновое распознавание

режим, позволяющий одновременно с распознаванием редактировать и сохранять распознанные страницы.

Ш

Шаблон блоков

в шаблоне описано положение и размеры блоков на странице.

Э

Эталон

набор пар “усредненное точечное изображение символа – его название”, который создается в процессе обучения системы на конкретном тексте.

Я

Яркость

параметр сканирования, определяет контрастность, т.е. различие между черными и белыми участками текста. При правильной настройке яркости качество распознавания увеличивается.

