

Практическая работа 153

Рассмотрим, как можно превратить бумажный текст в электронный с помощью программы FineReader (рис. 8.1).

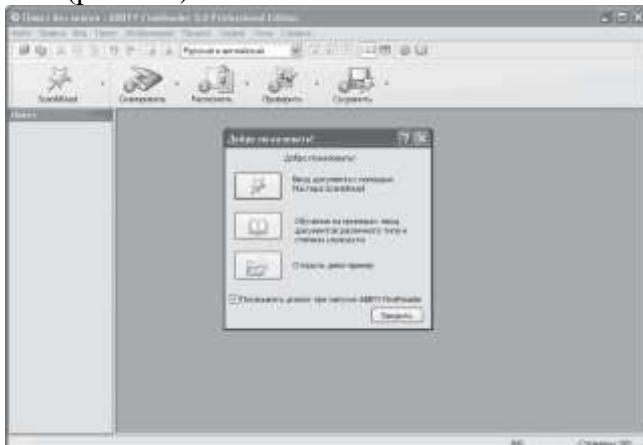


Рис. 8.1. Окно программы FineReader

Процесс сканирования в FineReader осуществляется двумя способами. Вы можете воспользоваться услугами Мастера Scan&Read, с помощью которого пройдете все четыре этапа преобразования документа бумажного вида в электронный (сканирование, распознавание, проверка и сохранение). Второй вариант – вручную пройти все эти шаги, выбирая соответствующие пункты меню либо используя кнопки панели инструментов.

После запуска FineReader и выбора режима работы программы (с помощью мастера или вручную) необходимо установить в сканер печатный документ. Для запуска процесса сканирования нажмите кнопку Сканировать либо выполните команду Файл > Сканировать изображение.

После выбора способа сканирования откроется окно, в котором можно выполнить предварительный просмотр и установить необходимые параметры (рис. 8.2). Это окно для разных типов сканера имеет разный вид, но все же основные его параметры одинаковы. Расскажу о наиболее общих параметрах сканирования на примере использования сканера Mustek 1200 UB Plus.



Рис. 8.2. Настройка параметров сканирования

Обратите внимание на то, как вы размещаете источник в сканере. Постарайтесь добиться, чтобы книга или журнал лежали как можно ровнее, ведь если текст расположить неровно, он будет распознан неправильно и вам придется вручную исправлять много ошибок.

Если вы имеете дело с текстом, следующим этапом вашей работы будет распознавание. Задача распознавания состоит в том, чтобы превратить отсканированное изображение в текст, сохранив при этом оформление страницы.

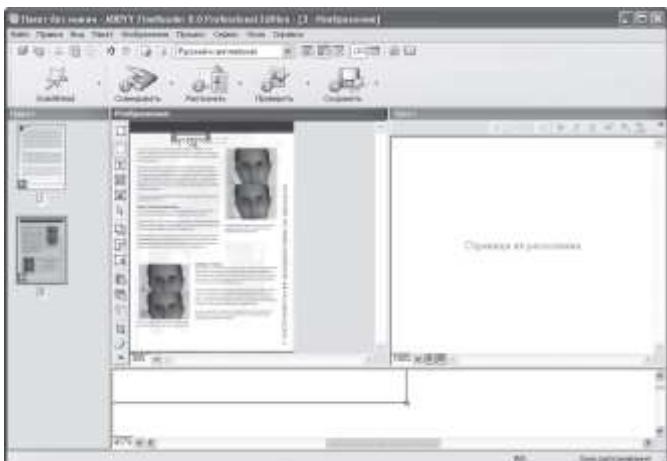


Рис. 8.3. Результат сканирования

Примечание

Если при сканировании вы сохранили результат в виде графического файла, его можно открыть для последующей обработки, выполнив команду **Файл > Открыть PDF-изображение**.

Для анализа макета страницы необходимо выполнить команду **Процесс > Распознать > Анализ макета страницы**. FineReader произведет автоматическое разбиение страницы на блоки (рис. 8.4). Для выделения или редактирования блока следует воспользоваться командой **Изображение > Изменить** тип блока и в появившемся меню выбрать нужный тип. Например, если у вас в тексте встречаются иллюстрации, пометьте их с помощью типа блока **Картинка** – выбрав пункт меню, вам следует выделить в окне **Изображение** нужный фрагмент. Точно так же помечаются текст и таблица.

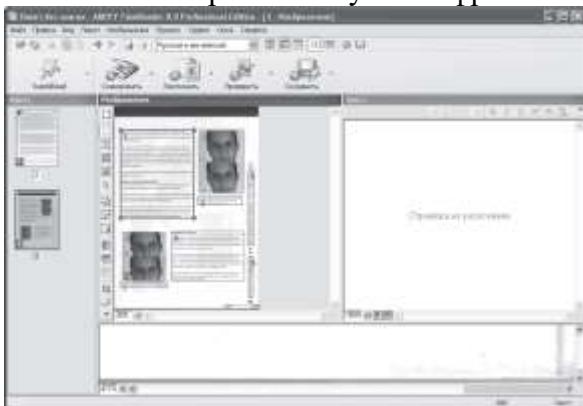


Рис. 8.4. Анализ макета страницы означает выделение на ней блоков разного типа

При обработке изображения, полученного в результате сканирования, FineReader выделяет блоки нескольких типов. Блок **Зона распознавания** используется для распознавания и автоматического анализа. В результате обработки он будет разделен на блоки других типов. Для корректного распознавания таблицы существует специальный блок **Таблица**, для распознавания текста – **Текст**, для изображений используются **Картинка** и **Штрих-код**.

Тип печати обычно определяется автоматически. Однако в некоторых случаях, особенно для текстов, напечатанных в черновом варианте или на матричном принтере, тип печати необходимо устанавливать вручную. Для этого выполните команду **Сервис > Опции**, перейдите на вкладку **Общие** и нажмите кнопку **Дополнительные опции**. В появившемся окне (рис. 8.5) выберите нужное значение с помощью переключателя **Тип печати**. По умолчанию здесь установлен переключатель **Авто**, но вы можете выбрать другой – например, **Пищущая машинка** или **Матричный принтер**.

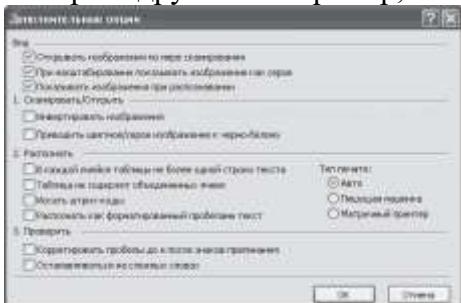


Рис. 8.5. Настройка параметров распознавания текста

После установки параметров можно начинать распознавание.

Результат распознавания будет отображаться в окне Текст, встроенном редакторе программы FineReader.

Примечание

Иногда программа по умолчанию неверно распознает блоки с вертикальным текстом. Для изменения ориентации текста щелкните правой кнопкой мыши на блоке с вертикальным текстом, выберите пункт Свойства и в открывшемся окне укажите нужный вариант в списке Направление текста. После этого еще раз распознайте этот блок.

Для проверки текста нажмите кнопку Проверить. На экране отобразится окно Проверка (рис. 8.6). В верхней его части система будет по очереди выделять найденные ошибки. Вы можете исправлять их непосредственно в этом окне. После исправления не забудьте нажать кнопку Подтвердить. В некоторых случаях программа FineReader будет предлагать варианты замены слова с ошибкой. Используя предложенные варианты исправления текста либо задав изменения вручную, можно исправить неверно распознанные слова.

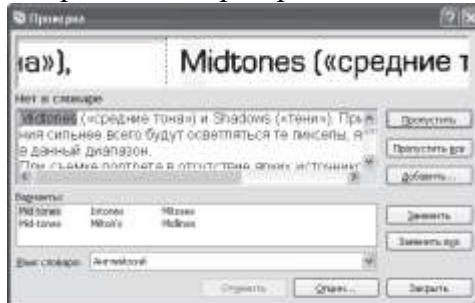


Рис. 8.6. После распознавания текст необходимо проверить

Во время проверки текста вы можете добавлять исправленные слова в словарь. Это позволит повысить качество распознавания, так как при распознавании система сверяется со словарем, в котором может не быть некоторых слов, особенно терминов или сокращений.

После завершения проверки закройте окно. Все исправления будут сохранены в распознанном тексте документа.

Получив готовый текст, вы можете его отформатировать – для этого предназначена панель инструментов Форматирование. На ней размещены инструменты для изменения шрифта и способа выравнивания текста.

После распознавания и исправления результаты работы можно сохранить в отдельном файле, скопировать в буфер обмена либо передать во внешнее приложение.

Один из способов сохранения результатов работы в FineReader – использование мастера сохранения результатов. Для его запуска нажмите кнопку Сохранить.

В окне мастера предлагается выбрать приложение для передачи текста либо одну из возможностей: Сохранить страницы, Отправить страницы по электронной почте, Копировать в буфер обмена (рис. 8.7). В этом же окне можно задать параметры сохранения оформления распознанного текста и возможность сохранения картинок.

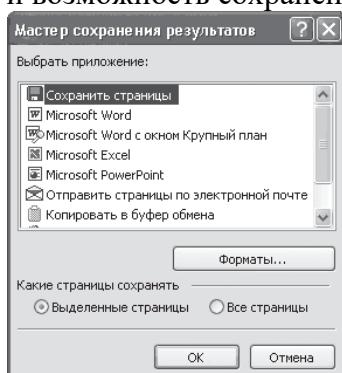


Рис. 8.7. Мастер сохранения результатов

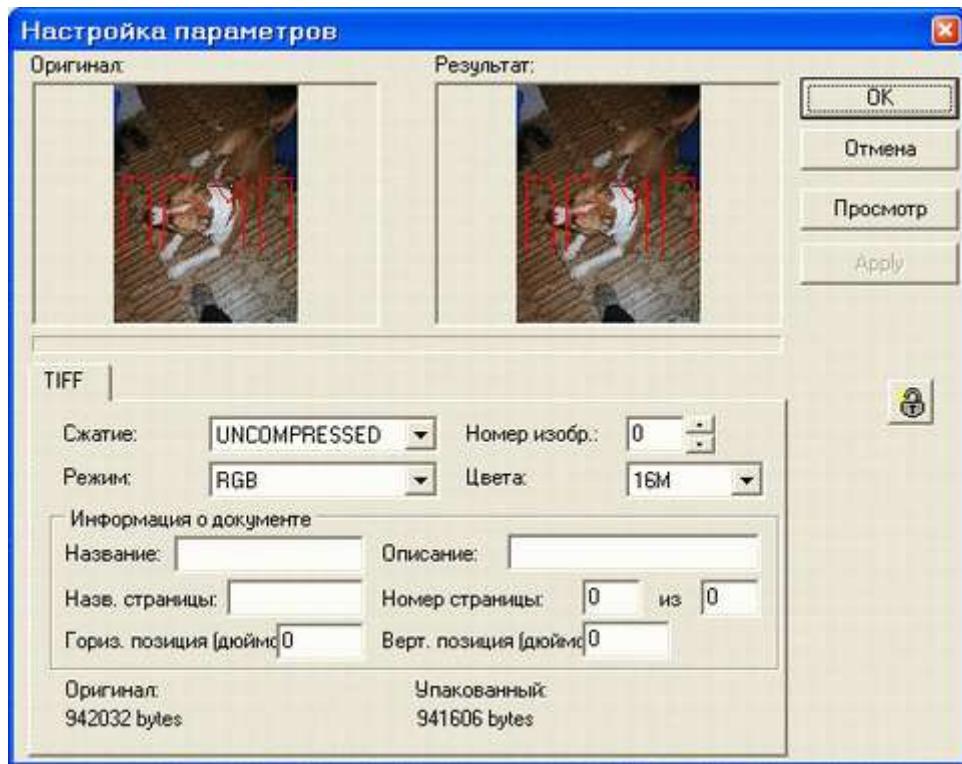
Выбрав пункт Microsoft Word, вы таким образом отобразите распознанный текст в окне текстового редактора Word.

После этого можно продолжить работу с документом.

Конвертация файлов

Конвертация файлов может производиться в пакетном и пофайловом режимах. Пофайловая конвертация осуществляется привычным для пользователей способом по команде Save as.

Откроется окно сохранения файла. Выбрав нужный формат, нажимаем "Сохранить". При необходимости сохранения файла в выбранном формате с параметрами, отличными от стоящих по умолчанию, нужно воспользоваться кнопкой "Дополнительно".



По этой команде появится окно настроек параметров сохранения изображения, в котором можно изменить параметры, свойственные каждому формату.

При необходимости пакетной обработки файлов порядок действий несколько иной. Для этого служит кнопка Batch Process. В появившемся после нажатия этой кнопки окне нужно составить список файлов для пакетной обработки, выбрать формат сохранения, при необходимости произвести изменение параметров формата, выбрать папку сохранения новых файлов и дать команду Start.

Форматы исходных файлов: DXF, TIFF, JPEG, PCX, BMP, DIB, RLE, PNG, WMF, EMF, TGA; TARGA, VDA, ICB, VST, PIX, PXM, PPM, PGM, PBM, ICO, CUR.

Форматы результирующих файлов: JPG, TIF, PCX, PNG, BMP, TGA, PXM, PPM, PGM, PBM.

Еще одна особенность программы: при конвертировании файлов можно изменять разрешение нового изображения и его размеры в пикселях.