



УЛУЧШЕНИЯ В ВОСЬМОЙ ВЕРСИИ ABBYY FineReader

- На 30% повышена точность распознавания факсов, на 15% — документов, отсканированных с низким разрешением (здесь и далее — по сравнению с версией 7.0, по результатам внутренних тестов ABBYY).
- На 40% повышена точность распознавания документов, снятых цифровым фотоаппаратом.
- Усовершенствован алгоритм обработки PDF-файлов, за счет чего время обработки сокращено вдвое.
- Добавлена возможность подключать при распознавании пользовательский словарь Microsoft Word. Это позволяет точнее распознавать тексты со специальной лексикой.
- Увеличилось количество поддерживаемых языков распознавания: всего их стало 179, для 36 из них предусмотрена проверка орфографии.
- Добавлена поддержка графического формата TIFF с LZW-компрессией.
- Добавлен новый формат сохранения: LIT (электронные книги Microsoft Reader eBook).
- При открытии многостраничных TIFF- и PDF-файлов FineReader теперь позволяет указать диапазон страниц, которые необходимо открыть.
- Появилась возможность создавать PDF-файлы, закрытые паролем.

Оцифровка ТЕКСТОВ

Предположим, вы любите печатать. Просто обожаете набирать тексты, нажимая клавишу за клавишей. Но даже в этом случае сомнительно, что необходимость загнать в компьютер стопятидесятистраничный реферат или доклад приведет вас в восторг.

Системой (или программой) распознавания текста называется приложение, основное назначение которого — переводить изображение (фотографию или результат сканирования) документа в редактируемый текстовый вид. Заметим, с полным и однозначным соответствием электронного варианта бумажному оригиналу, с корректно распознанными буквами, знака-

ми препинания и всей внутренней структурой документа. Именно эти задачи решает ABBYY FineReader всю свою долгую по меркам компьютерных программ жизнь — 12 лет. Секрет подобного долголетия прост; в любом офисе, в любом учреждении постоянно возникает необходимость сделать бумажный документ электронным, причем так, чтобы можно было его редактировать.

Даже в самом простом случае, когда документ состоит из одного лишь текста, без таблиц, иллюстраций и фона, «набивка» на клавиатуре требует времени — примерно 10–15 минут на каждую страницу. И как бы вы, проведя за подобной работой день-другой, восприняли предложение установить программу, которая сделает то же самое, но в 10–15 раз быстрее?

Долой рутину!

ABBYY FineReader позволяет избавиться от рутинной работы. Любой машинописный или печатный документ будет введен в компьютер с полным сохранением оформления документа (все иллюстрации, таблицы, списки). Очень полезным может оказаться умение программы ABBYY FineReader открывать PDF-файлы и также переводить их содержимое в редактируемый формат.

В общем-то, неважно, откуда оригинал: пришел по почте в виде PDF, получен со сканера в виде изображения или снят цифровым фотоаппаратом. В любом случае, после обработки с помощью ABBYY FineReader этот документ можно будет редактировать в программах Microsoft Office, сохранять в разных форматах, отправлять по электронной почте или публиковать в Интернете. Одним словом, это будет уже полноценная электронная редактируемая версия оригинала.

Для перевода документа в электронный вид нужно выполнить три несложных действия. Во-первых, нужно получить качественное изображение документа. Это можно сделать при помощи сканера или цифрового фотоаппарата. Если же речь идет об обработке PDF-файла, первое действие можно пропустить: файл уже содержит всю информацию, необходимую для работы системы. Во-вторых, изображение или PDF-документ следует загрузить в ABBYY FineReader и запустить распознавание. И в-третьих, нужно просмотреть (при необходимости — исправить) полученный электронный документ и сохранить его в желаемом формате.

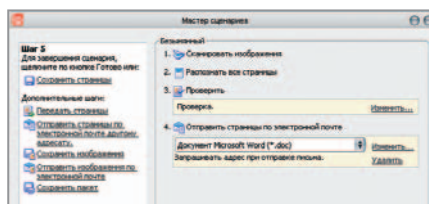
Экономия времени по сравнению с ручным вводом получается внушительная: примерно 14 минут на страницу. Другими словами, обработав упомянутый стопятидесятистраничный доклад, вы сэкономите примерно 35 часов — более четырех рабочих дней, с понедельника по пятницу включительно.

Работа со сценариями

Однако обычное распознавание — далеко не все, на что способна программа ABBYY FineReader 8.0. Эта версия (новейшая на сегодняшний день) может самостоятельно решать проблемы, ранее требовавшие участия человека. Например, FineReader определяет, получено изображение со сканера, с цифровой камеры или по факсу, и автоматически выбирает оптимальный способ обработки.

Да и саму обработку можно во многом переложить на «железные плечи компьютера». Допустим, требуется распознать пресловутые сто пятьдесят страниц. Для этого нужно сто пятьдесят раз нажать на кнопки «Сканировать» — «Распознать» — «Сохранить». Или один раз запустить сценарий обработки документов и заняться чем-нибудь поинтереснее, пока компьютер делает вашу работу.

«Научит» его этому, понятное дело, ABBYY FineReader 8.0, который позволяет работать автоматизированно. Для этого составляются так называемые сценарии — последовательности типовых операций, выполнение которых запускается одним нажатием кнопки. Предусмотрено несколько готовых сценариев (наподобие нашего «Сканировать» — «Распознать» — «Сохранить»), можно создавать и свои собственные.



↑ ABBYY FineReader позволяет автоматизировать работу

Впрочем, для эффективной работы с FineReader 8.0 не обязательно вникать в такие тонкости, как составление сценариев. Как и прежде, технология ABBYY под названием Scan&Read позволяет распознавать документы буквально нажатием одной кнопки.

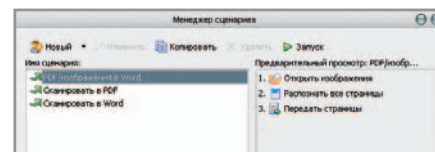
Как сделать лучше?

Конечно, определенные правила обработки документов все же существуют. Перечислим некоторые из них.

■ Сканировать желательно с разрешением не менее 300 точек на дюйм.

■ При работе с книгами (особенно толстыми) не стоит пользоваться CIS-сканерами: у этих тонких, компактных аппаратов небольшая глубина резкости. Поэтому текст, находящийся вблизи разворота (и потому не прижатый к стеклу сканера), может оказаться размытым до неузнаваемости.

■ Обращайте внимание на язык, на котором составлен документ. Ни одна система распознавания не способна справиться с текстом, если язык указан неверно. Даже несмотря на то что ABBYY FineReader «владеет» 179 языками и умеет распознавать документы, составленные на нескольких языках.



↑ Готовые сценарии отображаются в соответствующем Менеджере

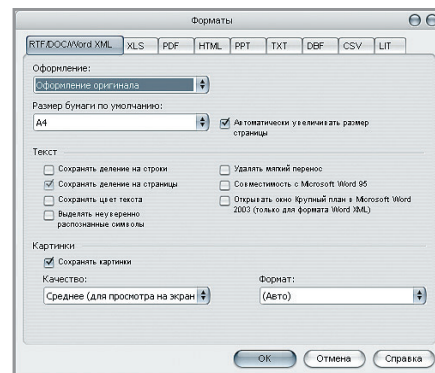
■ Если в документе нет цветных картинок, сканировать можно в полутоновом (сером) режиме, в противном случае предпочтителен цветной (24 бит) режим.

■ Если сканер и ABBYY FineReader по воле случая находятся на разных компьютерах, вы можете отсканировать материалы, сохранить их в популярном графическом формате и любым удобным способом «доставить» на тот ПК, который будет выполнять функции центра распознавания. Не стоит сохранять файлы в формате JPEG с большой степенью сжатия (компрессии). Мелкие детали изображения (к которым относятся и все буквы текста) при этом будут искажены, и общее качество распознавания понизится.

Фотоаппарат плюс FineReader

Но что делать в том случае, если сканера нет? То есть вообще нет нигде в пределах досягаемости? Никаких проблем — воспользуйтесь цифровым фотоаппаратом.

Это очень удобно, ведь цифровик намного компактнее и легче, не нуждается в питании от электрической сети, да и работает намного быстрее сканера. Конечно, цифровые снимки кое в чем уступают отсканированным изображениям. Например, на фотографиях документов освещение листа чаще всего неравномерное, какие-то области получаются ярче, другие темнее. Иногда, особенно при съемке с рук, страдает резкость. Фотоаппарат, в отличие от сканера, лишен крышки, прижимающей документ к оптиче-

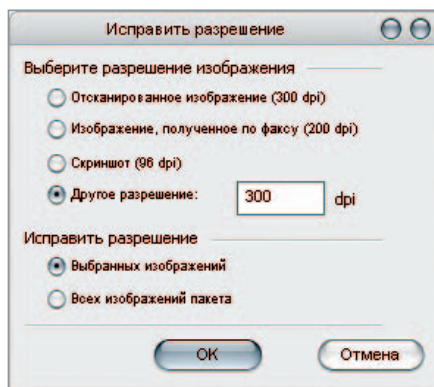


↑ Поддерживается работа с различными форматами текстовых документов



← **Возможна работа с PDF-файлами**

→ **Коррекция разрешения документа**



ской системе, поэтому нередко случаи искривления строк текста в кадре, особенно вблизи разворота книги.

Впрочем, ABBYY FineReader 8.0 умеет справляться со всеми перечисленными, а также со многими другими дефектами изображений. А чтобы облегчить ему задачу (и заодно ускорить обработку документов), рекомендуем при съемке для распознавания следовать нескольким простым правилам.

1. Используйте цифровой фотоаппарат с разрешением не менее 4 мегапикселей. Весьма желательно, чтобы он имел возможность отключения вспышки, ручную установку диафрагмы или режим приоритета диафрагмы, ручную или автоматическую фокусировку, функцию оптической стабилизации и объектив с переменным фокусным расстоянием (зум-объектив).
2. Перед съемкой убедитесь в том, что документ хорошо и равномерно освещен. Наилучшие результаты получаются при естественном освещении.
3. Отключите вспышку (в большинстве любительских цифровиков по умолчанию установлен автоматический режим вспышки).

4. Установите режим макросъемки (Close-up или Macro). Обычно этот режим обозначается пиктограммой в виде цветка.

5. Расположите фотоаппарат параллельно плоскости документа, приблизительно в центре страницы.

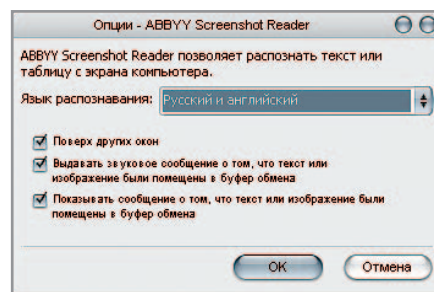
На карту памяти объема 128 Мбайт помещается около сотни снимков в разрешении, достаточном для уверенной работы ABBYY FineReader 8.0. Учитывая тот факт, что при распознавании одного документа (по сравнению с ручным набором) вы экономите примерно 14 минут, нетрудно подсчитать, что одна флэшка, заполненная снимками для распознавания, освобождает вам 24 часа. Сутки, которые не пришлось провести за компьютером. И которые можно потратить на что-нибудь более увлекательное! Например, на съемку окрестных красот — с помощью все того же цифровика.

Ничего невозможного

Самым, пожалуй, оригинальным нововведением восьмой версии является ABBYY Screenshot Reader. Это простое и удобное

приложение предназначено для распознавания текста из любой области экрана компьютера. Screenshot Reader позволит переводить в редактируемый формат такие тексты, которые нельзя скопировать обычным способом (например, тексты с интернет-страниц, созданных при помощи технологии Flash). Как распознанный текст, так и снимок экрана могут быть сохранены в виде файла или переданы в буфер обмена. Таким образом, при помощи ABBYY Screenshot Reader можно:

- быстро собрать в один файл цитаты из нескольких открытых документов;
- получить в виде текста содержание папки с файлами, названия файлов и атрибуты;
- сохранить историю писем в Outlook;
- извлечь текст из сообщений об ошибках, сообщений в строке статуса;
- сохранить данные только из выбранного столбца таблицы.



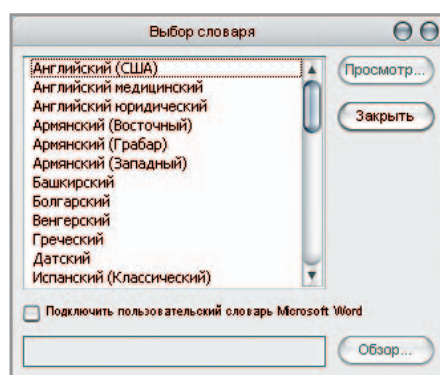
Окно настройки ABBYY Screenshot Reader — простое и понятное

Утилита доступна зарегистрированным пользователям ABBYY FineReader 8.0. Работа с ней настолько проста, что не требует обучения: достаточно просто указать, в какое именно приложение следует передать распознанный текст. **Сергей Михайличевский**

НОВЫЕ ВОЗМОЖНОСТИ ABBYY FINEREADER 8.0

- Распознавание цифровых фотографий текстовых документов.
- Автоматическая проверка разрешения входных изображений.
- Функция обрезки изображения, ручной и автоматический режим. Предназначена для удаления черных полей, приведения страниц пакета к одинаковому размеру, для удаления с фотографий документов областей, не содержащих текста.
- Функция выпрямления искривленных строк, получающихся при сканировании и фотографировании толстых книг и подобных документов.
- Распознавание снимков экрана. Зарегистрированным пользователям доступна утилита ABBYY Screenshot Reader.

- Автоматизированная обработка документов — как по типовым сценариям приложения, так и по сценариям, созданным пользователем.
- Новый режим распознавания — «Быстрый». Он позволяет в 2–2,5 раза ускорить обработку документов с простыми оформлением и структурой.
- Распознавание и сохранение гиперссылок в итоговом документе.
- Прямая, без предварительного сохранения на диск, передача PDF-документов в Adobe Acrobat или Adobe Reader.
- Создание PDF-файлов с тегами. Такие файлы удобно просматривать на экранах любого размера, в частности, на экранах карманных компьютеров.



Возможно корректное распознавание текстов на 179 языках