

Интеллектуальное распознавание текстов

Непревзойденная точность преобразования документов и PDF-файлов



FineReader 9: Свежим взглядом

ТОЧНОСТЬ

Автоматически определяет язык, сохраняет структуру документа

reconnaissance OCR and PDF conversion за минимальное время
 exact format retention Layout, Übersetzung
 максимальная точность



ABBYY®

FineReader® Professional Edition OCR 9.0

СКОРОСТЬ

Интеграция с офисными приложениями, сценарии типичных операций

Можете освоить десятипальцевый метод набора на клавиатуре или нанять секретаршу — в любом случае FineReader 9.0 оцифрует «бумажные» документы быстрее

В этой статье

Новые функции FineReader 9.0

Адаптивное распознавание

Сценарии типичных операций

Точность распознавания: экспресс-тест

Система распознавания текстов ABBYY FineReader относится к тем немногим программам, у которых практически нет конкурентов. С одной стороны, это не очень хорошо: ведь здоровая конкуренция, как известно, всегда идет на пользу и производителям, и потребителям. Но, с другой стороны, такая ситуация вызвана прежде всего тем, что в основу FineReader заложены достаточно качественные алгоритмы распозна-

вания текстовых документов, в том числе на разных языках и с различными элементами оформления.

НОВОВВЕДЕНИЯ

Обработка текста ускоряется

К 179 естественным и искусственным языкам, которые поддерживались в восьмой версии FineReader, не добавилось ни одного нового. Но зато теперь уже не нужно перед запуском распознавания выбирать из списка нужный язык (что пользователи частенько забывают сделать) — программа определит его сама. К тому же благодаря этому новый FineReader справляется с текстами, где на одной странице встречаются разные языки.

СНП Вывод

Компания ABBYY взяла курс на улучшение уже имеющихся возможностей своего продукта — и это вполне понятно: ведь в системе распознавания текстов FineReader уже есть все необходимые функции. Улучшенные средства распознавания сносок, колонтитулов и других сложных элементов, автоматическое определение языка — для пользователя все это хороший повод задуматься о переходе на новую версию. Не стоит забывать и о том, что ABBYY FineReader 9 работает в Windows Vista, поддерживает новые форматы Microsoft Office 2007, а также использует возможности многоядерных процессоров.



Выбирать изображения для распознавания тоже стало проще: в окне открытия файла появилась область предпросмотра — впрочем, если файл слишком большой и долго открывается, ее можно отключить.

Расширилась и совместимость FineReader: если восьмая версия запускалась только из MS Word, то кнопка вызова девятой появилась в Excel и Outlook, а также в «Проводнике».

ИНТЕРФЕЙС

Удобное рабочее место

Обновленный интерфейс FineReader оценят не только пользователи, много работающие с этой программой, но и те, кто обращается к ней время от времени. Основные изменения затронули инструменты для работы с изображениями и редактирования распознанного текста. Если раньше команда проверки орфографии размещалась на панели «Scan&Read», то теперь она переместилась в окно для работы с текстом, так что пользователь видит ее только тогда, когда она действительно нужна. Инструменты редактирования графики переселились из меню «Изображение» в новое окно, которое так и называется — «Редактор изображений».

В нижней части блоков появилась панель свойств, содержащая дополнительные инструменты, в зависимости

от выполняемой операции. Например, при редактировании текста на этой панели появляются инструменты для выбора стилей, изменения регистра букв, установки цвета фона, верхнего и нижнего индексов.

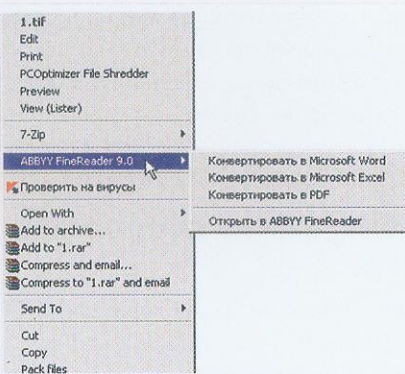
Панели инструментов, которые прежде отображались по умолчанию, в FineReader 9 стали не нужны, так что в окне программы стало больше полезного места. Вместо нескольких панелей инструментов появилась одна удобная боковая панель, на которой размещаются кнопки для быстрого сканирования, открытия, создания и сохранения документа, настройки программы, выбора языков. Здесь же можно увидеть все страницы документа и ход распознавания. Боковая панель может быть скрыта с экрана одним кликом мыши, что еще больше расширяет рабочую область.

МАСТЕР СЦЕНАРИЕВ

О рутине позаботится компьютер

Пользователи, не очень хорошо знакомые с FineReader, наверняка оценят новый мастер сценариев. Впервые сценарии были представлены в восьмой версии FineReader, а в девятой они стали еще удобнее.

Мастер сценариев теперь появляется сразу после запуска программы; к нему также можно в любой момент вернуться, кликнув по кнопке «Основные сценарии». Кроме того, сценарии FineReader 9 вызываются прямо из меню «Пуск» операционной системы.



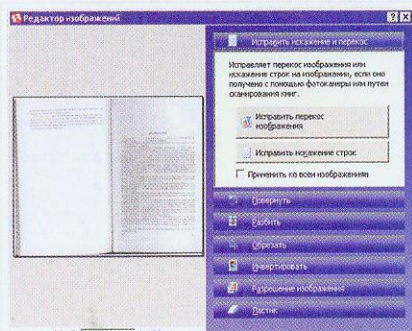
ЭКСПОРТ Один из сценариев запускается прямо из файлового менеджера

Получить результат за одно нажатие



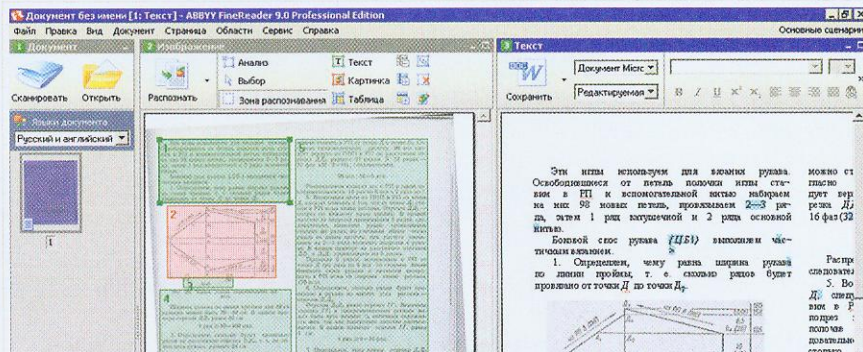
МАСТЕР СЦЕНАРИЕВ поможет быстро выполнять типичные операции

Например, если выбрать сценарий «PDF или изображения в Microsoft Word», то FineReader откроет окно для выбора файла и сразу начнет его сканировать, а потом создаст новый документ Word с распознанным текстом. Отметим также, что один из сценариев распознает документы, снятые цифровой камерой, — иногда снять →



РЕДАКТОР ИЗОБРАЖЕНИЙ

В новом окне собраны все средства для работы с графикой



НОВЫЙ ИНТЕРФЕЙС Рабочая область увеличена: вместо панелей инструментов теперь все нужные функции собраны на боковой панели

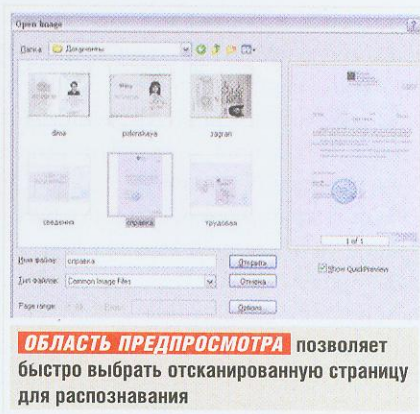
текст цифровиком проще и быстрее, чем сканировать, — ABBYY рекомендует в этом случае использовать камеру с разрешением не менее 4 мпикс.

РАЗВИТИЕ ТЕХНОЛОГИИ
Сохранение оформления

Одно из ключевых нововведений FineReader 9 — адаптивное распознавание многостраничных документов (см. CHIP № 11, 2007, «OCR: от распознавания символов к воссозданию структуры документа»). Чтобы оценить преимущества новой технологии, получившей название ADRT (Adaptive Document Recognition Technology), достаточно распознать документ с большой таблицей. В преды-

дущих версиях приходилось объединять такие таблицы вручную, поскольку FineReader располагал их на разных страницах — так же, как в оригинале. То же самое получалось и с абзацами, которые переходили с одной страницы на другую. FineReader 9 определяет, что такой абзац или таблица — единый объект, и автоматически «склеивает» их части.

Такие элементы как сноски, колоннотитулы, подрисуночные подписи и номера страниц благодаря технологии ADRT преобразуются в соответствующие объекты Word. Например, при удалении сноски из середины распознанного документа номера следующих сносок автоматически корректируются, а отредактированный колоннотитул изменяется на всех страницах.



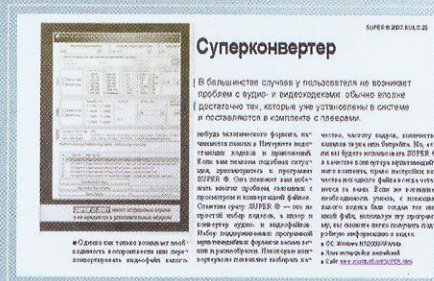
ОБЛАСТЬ ПРЕДПРОСМОТРА позволяет быстро выбрать отсканированную страницу для распознавания

«Умное» распознавание затронуло и PDF-документы: если они содержат текстовый слой, FineReader 9 извлекает текст без распознавания, что ускоряет обработку распознанного текста.

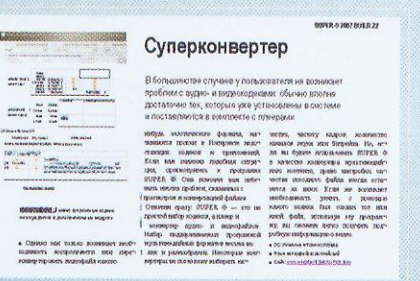
■ ■ ■ Сергей и Марина Бондаренко

Экспресс-тест: точность распознавания

Наверное, у каждого, кто использует FineReader, есть тайная мечта — чтобы точность распознавания текста и оформления документов однажды достигла ста процентов. Очень может быть, что это когда-нибудь произойдет, однако в девятой версии разработчикам этого достичь пока что не удалось. Согласно официальной информации, основанной на результатах внутреннего тестирования ABBYY, в FineReader 9 значительно повышена точность распознавания договоров (на 19%), книг (на 22%), журналов и газет (на 32%). У нас нет повода сомневаться в этих данных, однако проведенное нами сравнение восьмой и девятой версий дало неожиданные результаты. Мы отсканировали страницу журнала CHIP, на которой, кроме текста, было еще две иллюстрации, и поочередно распознали ее в FineReader версиями 8 и 9. Восьмая версия безошибочно определила графические блоки, а вот девятая, к нашему удивлению, распознала одну из иллюстраций как текст. Справедливости ради отметим, что эта иллюстрация представляла собой скриншот, на котором явно различимы текстовые элементы. Вероятно, именно повышенная точность распознавания и стала причиной ошибки в этом случае. Целью второго теста было определение того, насколько хорошо новый FineReader справляется



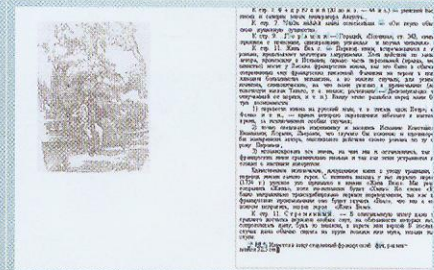
ЧРЕЗМЕРНОЕ УСЕРДИЕ... FineReader 8 (слева) безошибочно определил графический элемент, в то время как «девятка» (справа) попыталась вычленил из него текст



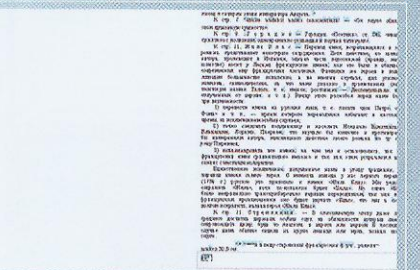
ЧРЕЗМЕРНОЕ УСЕРДИЕ... FineReader 9 (справа) попыталась вычленил из него текст

со страницами книг. Мы отсканировали разворот старой книги, на которой, кроме текста, была еще иллюстрация. FineReader 8 хоть и помнил некоторые текстовые символы как сомнительные, однако блок картинки определил безошибочно. А вот результат, выданный версией 9, снова оказался неожиданным: про-

грамма вообще «не заметила» иллюстрацию, и на готовом документе на ее месте зияла дыра. Правда, скорость распознавания в FineReader 9 выше, чем у версии 8. Отметим также, что выявленные в обоих текстах недостатки легко устраняются — достаточно лишь выделить картинку вручную.



...И НЕДОСТАТОЧНАЯ РАЗБОРЧИВОСТЬ Девятая версия FineReader (справа) вообще отказалась признавать существование малоконтрастной картинки



...И НЕДОСТАТОЧНАЯ РАЗБОРЧИВОСТЬ Девятая версия FineReader (справа) вообще отказалась признавать существование малоконтрастной картинки