

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Адаптивные технологии FineReader 9.0

ДМИТРИЙ НИКИШИН

МНОГО ЛЕТ ПРОИЗВОДИТЕЛИ OCR-программ стремились добиться качественного распознавания символов, слов, фраз и целых страниц. Сейчас эта задача уже в основном решена, и перед разработчиками стоит другая цель: обеспечить распознавание документа как единого, целостного объекта, сведя к минимуму ручное редактирование.

OCR-СИСТЕМЫ

Главными новшествами FineReader 9.0 являются технология адаптивного распознавания ADRT (Adaptive Document Recognition Technology) и функция распознавания документов с фотографий. Новый FineReader обрабатывает многостраничные документы не страница за страницей, а как единое целое. Благодаря этому происходит корректное распознавание таких элементов, как колоннотитулы, сноски, шрифты, стили, нумерация страниц, примечания и нумерованные подписи к таблицам и изображениям, сохраняется целостность абзацев, разбитых на две страницы. Система FineReader 9.0 способна самостоятельно находить соответствие между использованными в исходном документе и установленными на компьютере пользователя шрифтами.

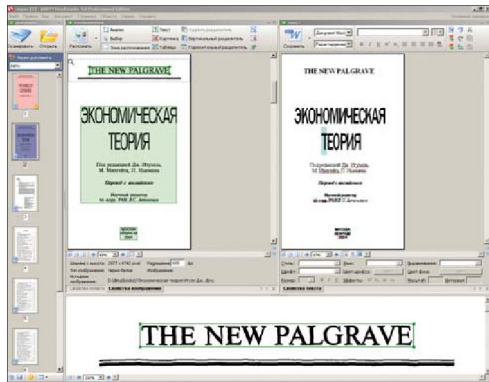
В новой версии ABYY FineReader разработчик увеличил число поддерживаемых языков до 179. Программа способна самостоятельно определить используемые в документе языки, предусмотрена также возможность их ручного выбора.

FineReader может обрабатывать как отсканированные копии, так и фотографические снимки документов. Программа неплохо справляется с обработкой снимков, в том числе недо-

статочно качественных. В руководстве пользователя приведены подробные советы по фотографированию документов для их последующего распознавания.

По сравнению с предыдущими версиями интерфейс программы стал проще и понятнее. Для выполнения наиболее распространенных задач, таких как получение документа в форматах Word или PDF, предусмотрены сценарии «в один клик». Пользователи могут создавать собственные сценарии, описывающие последовательности действий

Девятая версия программы вышла в трёх релизах: Home Edition — для домашних пользователей и учебных заведений, Professional Edition — для профессионального использования и Corporate Edition — для работы в корпоративной сети. FineReader Corporate Edition поддерживает автоматическую пакетную обработку изображений и PDF-документов: исходные файлы могут поступать в папку на локальном диске, на другой машине, которая доступна по сети, на ftp-сервере или в папку почтового ящика. Все опе-



Технология ADRT позволяет распознать документ, сохраняя шрифты и стили оригинала



Панель для выбора основных сценариев

рации выполняются в соответствии с описанием. Можно, например, назначить распознавание всех поступивших за день файлов на ночное время, когда компьютер не выполняет других задач.

Благодаря интеграции с Microsoft Office 2007 документы можно теперь распознавать непосредственно из Microsoft Word, Microsoft Excel и Microsoft Outlook, сохраняя результат в форматах DOCX и XLSX. Кроме того, появилась возможность сохранять документы в формате PDF/A, применяемом для долгосрочного хранения документов в архивах и библиотеках.

Полезной возможностью новой версии ABYY FineReader стала поддержка распознавания текста из файлов популярного формата DjVu, который применяется для хранения книг и документов. Это позволяет непосредственно

обрабатывать DjVu-файлы, не переводя их предварительно в другие графические форматы.

Для повышения производительности в новую версию FineReader добавлено распараллеливание задач на современных многоядерных процессорах. Это ускоряет распознавание больших документов. Кроме того, пользователь может редактировать уже обработанный текст, в то время как программа продолжает распознавание в фоновом режиме.

Юридические и бухгалтерские документы обычно содержат подписи, которые важно сохранить в неизменном виде после распознавания. В предыдущих версиях программы FineReader подписи необходимо было выделять вручную. Новая версия обрабатывает подписи автоматически, сохраняя их в электронной копии в виде графического объекта.

Программа ABYY FineReader 9.0 официально сертифицирована для устройств и программного обеспечения Windows Vista. Гарантирована совместимость программы с дополнительными функциями ОС Microsoft Windows Vista.

В новой версии ABYY FineReader разработчик сделал акцент на автоматизации и уменьшении ручных операций. Технология ADRT значительно ускоряет распознавание сложных документов, повысив при этом качество обработанных копий.

Пользовательские сценарии позволяют даже не знакомому с FineReader человеку легко начать работу и быстро получить результат. Благодаря поддержке многоядерных процессоров возросла скорость обработки документов. Новая версия ABYY FineReader найдёт применение как у конечных пользователей, так и в крупных компаниях, которые часто сталкиваются с необходимостью перевода документов из печатного в электронный вид.

Обновление Vista всё же выйдет!

АНДРЕЙ КОЛЕСОВ

4 ФЕВРАЛЯ КОРПОРАЦИЯ Microsoft объявила о запуске в производство новой серверной ОС Windows Server 2008 и первого сервисного пакета Windows Vista Service Pack 1 (SP1). Казалось бы, две новости совершенно разного калибра — в первом случае речь идет о начале смены самой популярной платформы корпоративного уровня, во втором — вроде бы о сугубо техническом процессе обновления выпущенного год назад продукта. Но реакция СМИ и интернет-общественности показывает: именно информация о Vista SP1 вызвала наибольший интерес. И это вполне понятно.

Напомним, что Vista и Server 2008 — это не просто очередные версии знакомых продуктов, они знаменуют смену поколения ОС Microsoft, при которой, помимо многих других новшеств, фактически происходит замена архитектуры ядра NT (имеющей уже почти 20-летнюю историю) на ядро Longhorn. По сути, год назад именно с выпуском Vista начался этап глобальной опытной эксплуатации Longhorn, что было понятно всем ИТ-профессионалам, да и Microsoft не особенно скрывала это, подчеркивая, например, что она не включила сразу в Vista ряд технологий, которые будут доступны позднее, с выходом серверного варианта. Сейчас период «опытной эксплуатации» заканчивается.

Вообще-то разговоры об обновлении Vista пошли еще даже до ее выпуска. Речь тогда шла о появлении пакета обновлений для Vista уже через несколько месяцев после ее официального выхода на рынок, а затем о выпуске, спустя год (т. е. к настоящему времени) Vista R2 (тогда она имела кодовое название Vienna). Оценивая сейчас ту информацию, можно предположить, что это был лишь маркетинговый ход Microsoft, чтобы успокоить общественное мнение, взбудораженное широко обсуждавшимся вопросом, когда следует переходить на новую ОС.

Впрочем, корпоративные заказчики в таких дискуссиях и не участвовали — они давно и отлично знают, что нормальная работа с принципиально новым продуктом начинается только с выхода его третьей версии, а с обновленной версией — после появления сервисного пакета, причем лучше не первого, а второго. Так или иначе, но лишь прошлым летом Microsoft публично заявила о начале подготовки Vista SP1, сказав, что он появится, возможно, осенью. Но аналитики сразу дали другой прогноз — пакет обновлений выйдет одновременно с серверной ОС в феврале. Примерно так и получилось со смещением еще на месяц: Vista SP1 для шести языков первой приоритетной группы выйдет в марте, русскоязычный вариант ожидается в апреле.

Сроки выхода SP1 — это, конечно, интересно, но гораздо важнее — какие же новшества ожидают пользовате-

лей и ИТ-специалистов? Надо сказать, что в этом вопросе до сих пор ясности нет. В пакете будут лишь исправлены ошибки или все же появятся и какой-то дополнительный функционал? Будет ли, например, исправлен довольно очевид-

Vista SP1 для шести языков

первой приоритетной группы

выйдет в марте, русскоязычный

вариант ожидается в апреле.

ный дефект при работе с Kontakтами Windows, который Microsoft почему-то объявила «фичей» (специфической особенностью продукта)? Последние месяцы представители корпорации явно избегали обсуждений подобных вопросов, ссылаясь на то, что конкретный состав SP1 будет определен лишь в последний момент. Этот момент настал — началось производство дисков с обновлениями.

Официальная информация о Vista SP1 выглядит сейчас так:

- улучшена поддержка устройств: сертифицированных для Vista компонентов и устройств более чем от 600 производителей, Vista SP1 поддерживает около 80 тыс. устройств и компонентов, что почти в два раза больше, чем на момент выхода ОС. Это означает, что свыше 40 тыс. уст-

ройств, не совместимых с Vista на момент ее выхода, теперь работают корректно;

- улучшена совместимость Vista с продуктами сторонних производителей: Vista SP1 поддерживает в 10 раз больше приложений для пользователей и малого бизнеса, чем на момент выпуска, и теперь совместима со 150 главными бизнес-приложениями от таких производителей, как Symantec, Nortel, Cisco, SAP, Adobe, Citrix, Oracle, Sun и IBM;

- увеличено время функционирования от батарей при работе на портативном ПК в среднем на 7% и до 10% в отдельных случаях;

- повышена стабильность работы: устранено большинство причин сбоев ОС и ПО Microsoft, при этом среднее время между отказами системы увеличилось более чем в два раза, по сравнению с первоначальным выпуском Windows Vista;

- улучшена производительность: устранены причины задержек при копировании файлов на ПК, при этом копирование в отдельных случаях происходит до 50% быстрее;

- уменьшилось время, необходимое на выход ноутбука из «спящего» режима;

- пакет включает в себя все обновления безопасности, появившиеся с момента выхода Windows Vista. При этом Microsoft подчеркивает, что за первый год после появления Windows Vista в ней было обнаружено в два раза меньше уязвимостей, чем за первый год после выхода таких ОС, как MacOS 10.4, RedHat Linux, Ubuntu и Windows XP.

Более точную информацию мы сможем получить через один-два месяца, познакомившись лично с Windows Vista SP1.