



# С бумагой

# на экран

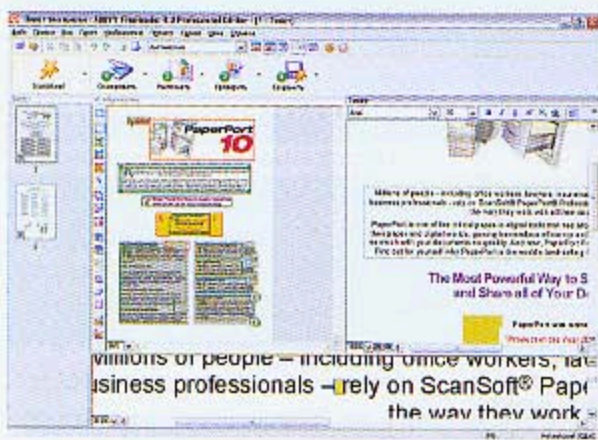
...и, конечно, не забыть про безопасность. В настоящее время в России нет единой системы сертификации компьютерной техники, поэтому при покупке необходимо обращать внимание на наличие сертификата соответствия. Кроме того, важно выбирать продукцию известных производителей, чтобы избежать проблем с качеством и надежностью. В заключение хочется отметить, что рынок компьютерной техники в России продолжает развиваться, и в ближайшее время можно ожидать появления новых, более совершенных моделей устройств. Это позволит пользователям получать еще более качественный и удобный опыт работы с компьютером.

Перевести документ из компьютера на бумагу – задача принтера. А обратно, с бумаги в компьютер, – сканера и специального программного обеспечения – OCR-системы

СИСТЕМЫ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ символов (Optical Character Recognition – OCR) предназначены для автоматического ввода в компьютер печатных документов. Изначально изображение, полученное со сканера или цифровой фотокамеры, – это не редактируемый текст, а всего лишь растровое изображение. OCR-система делит эту картинку на фрагменты, выделяет в них символы и распознает их. Мы решили сравнить функциональность и скорость работы наиболее распространенных у нас OCR-систем, поддерживающих кириллицу, – ABBYY FineReader 8.0 Professional и OmniPage 15 Professional.

## OCR-ПОЛИГЛОТ

Каждая следующая версия FineReader быстрее распознает все бо-

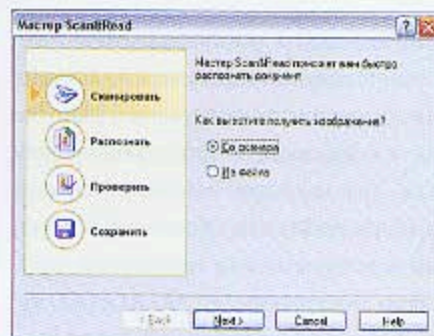


FineReader 8.0 – последняя версия популярной OCR-системы разработки ABBYY

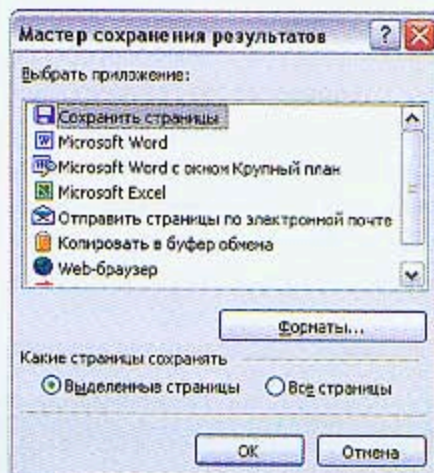
лее сложные тексты. В последней, восьмой, версии «порог чувствительности» снизился до разрешения 200 dpi. Благодаря этому появилась возможность, во-первых, ускорить сканирование документов, а во-вторых, распознавать тексты, снятые цифровой фотокамерой с разрешением от пяти мегапикселей. Другими словами, можно, например, прийти в библиотеку, отсканировать нужные страницы и пойти домой, минуя очередь на ксерокопирование.

Из остальных функций, появившихся в FineReader 8, следует отметить интеграцию в Microsoft Office, дополнительную утилиту для распознавания копий экрана Screenshot Reader и встроенную обучающую систему для начинающих пользователей.

Установка FineReader максимально автоматизирована. В стандартном режиме от пользователя требуется лишь выбрать папку, в которую будет установлена программа, и язык интерфейса. При выборочной установке можно отказаться от инсталляции компонентов, которые вам не нужны. Последний этап установки FineReader – активация



Для автоматизации сканирования и распознавания документов в FineReader используется мастер Scan&Read



Документ, распознанный FineReader, можно сохранить в различных форматах

системы. Это можно сделать с помощью дискеты, входящей в комплект поставки, по электронной почте или через интернет. Еще несколько лет назад наиболее распространенным был первый вариант, однако сейчас флоппи-дисководы постепенно уходят в прошлое, и пользователям все чаще приходится активировать продукты по сети.

Интерфейс FineReader хорошо продуман с точки зрения эргономики. Возможность переключения системы меню на русский и украинский языки выгодно отличает программу от американского конкурента. Кнопки основных режимов работы с большими пиктограммами и подписями вынесены на отдельную панель. Первая из этих кнопок, Scan&Read, запускает мастер ска-

## Точки и полосы

Чем выше качество изображения – то ли отпечатанного, то ли экранного, – тем проще распознавать содержащийся в нем текст. Численно этот параметр принято оценивать количеством точек на дюйм (dots per inch – dpi): поскольку компьютерные мониторы и цифровые печатающие устройства разбивают изображение на строки, а каждую строку – на точки, то чем больше таких точек при-

ходит на единицу длины (по традиции – дюйм), тем выше качество изображения. Для отпечатанных изображений стандартным считается качество 300 dpi. К экранной графике требования значительно ниже – 75 dpi. Таким образом, для распознавания текста из копий экрана требуются более сложные программные алгоритмы, чем для распознавания даже плохо отпечатанного текста.

нирования и распознавания текстов. С его помощью можно быстро отсканировать, открыть, распознать и проверить орфографию текста. Затем другой мастер, сохранения результатов, помогает сохранить документ или отдельные страницы в формате doc, xls или pdf, отправить по электронной почте или скопировать в буфер обмена. Таким образом, распознавание страницы сводится буквально к нескольким щелчкам мыши.

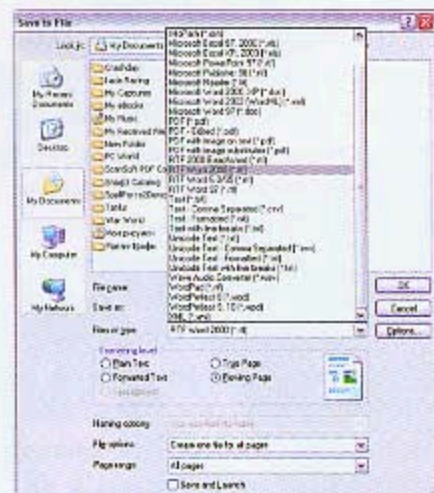
### ФОРМАТ НА ЛЮБОЙ ВКУС

В отношении разрешения распознаваемых изображений OCR-система OmniPage 15 производства компании Nuance Software несколько

шим количеством ошибок, которые придется исправлять вручную.

Установка программы стандартная: от пользователя требуется выбрать папку и способ установки – обычный либо выборочный. При первом запуске OmniPage нужно активировать через интернет. Если зарегистрироваться в Сети, то можно будет покупать следующие версии программы со значительной скидкой.

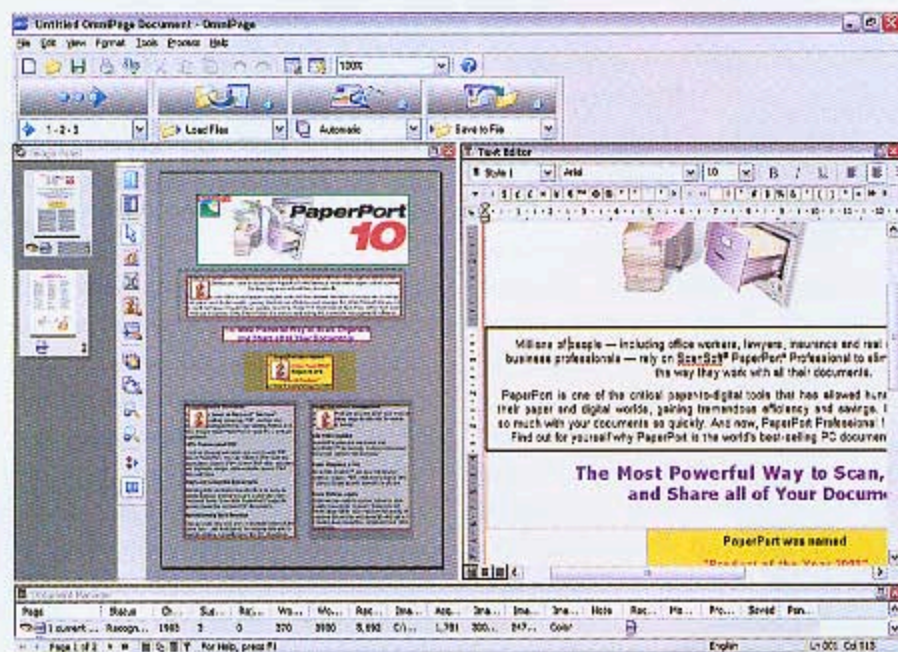
По сравнению с предыдущими версиями, интерфейс OmniPage 15 стал гораздо приятнее на вид. Правда, его языком по-прежнему остается английский. Язык распознаваемого текста выбирается не на панели инструментов, как в FineReader, а в настройках программы.



Документ, распознанный в OmniPage, сохраняется во множестве форматов

матически фрагментирует и распознает изображение, используя заранее сделанные настройки. Распознанный текст можно сохранить в различных стандартных форматах: html, разных версиях Microsoft Word и Excel, WordPerfect, xml, pdf и др. В целом, все просто, разве что перед сканированием и распознаванием приходится серьезно работать с предварительной настройкой программы.

В комплекте с OmniPage поставляется несколько полезных утилит. В частности, PDF Converter позволяет быстро и удобно преобразовывать файлы формата PDF в редактируемые форматы: doc, rtf, wpd, xls. Упрощенный вариант утилиты PDF Create! выполняет обратное преобразование: превращает практически любой текстовый или графический файл в формат PDF.



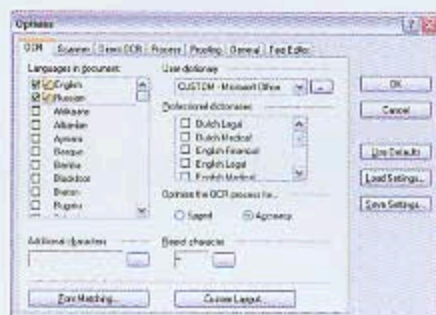
OmniPage 15 – еще одна распространенная у нас OCR-система, поддерживающая кириллицу

требовательнее, чем FineReader: минимально допустимое для нее разрешение составляет 300 dpi, а рекомендуемое – 400 dpi. При сканировании документа с минимальным разрешением есть риск получить недостоверный текст с боль-

шоем количеством ошибок, которые придется исправлять вручную. Кнопки основных функций здесь тоже выполнены в виде больших пиктограмм, под которыми располагаются меню со списками стандартных возможностей. Мастера сканирования и распознавания в OmniPage нет – программа авто-

### СКОРОСТЬ И КАЧЕСТВО РАСПОЗНАВАНИЯ

Сравнивая OCR-системы, мы обращали внимание не только на скорость, но и на качество распознавания текстов. Чем меньше ошибок в распознанном тексте, тем меньше времени уйдет на доработку доку-



Выбор языка распознавания и настройка других параметров в OmniPage

мента. Для тестирования использовалось несколько документов на русском языке и один на английском.

Как видно из таблицы, FineReader 8.0 обеспечивает высокое качество и достаточную скорость распознавания различных документов. OmniPage, будучи пакетом бельгийской

разработки, в документах, написанных на русском языке, допустил хоть и не много, но все же больше ошибок, чем FineReader. По времени распознавания русского текста OmniPage также несколько отстает. Зато эта программа заметно превосходит FineReader в распознавании англоязычных документов как по качеству, так и по скорости.

В целом, OmniPage лучше даются комбинированные изображения с большим количеством изображений. Программа эффективнее отделяет текст от графики. FineReader же, напротив, лучше работает с чистым текстом. Впрочем, это касается скорости, а отнюдь не качества.

Итак, для того чтобы система распознавания хорошо работала с тек-

стом на определенном языке, она должна создаваться людьми, хорошо знакомыми с этим языком. Для распознавания документов, написанных на русском, наиболее подходящим выбором будет FineReader. OmniPage в этом отношении немного, но все же уступает, хотя с документами, написанными на английском языке, эта система справляется без проблем. Кроме того, нельзя не учитывать возможность быстрой настройки FineReader на распознавание множества других языков (в том числе и украинского), а также пирокую доступность у нас этого пакета. Таким образом, FineReader по-прежнему остается лидирующей OCR-системой.

Александр Гусленко  
ag@comizdat.com

## OCR-системы: основные характеристики и качество работы

OCR-система	OmniPage 15	FineReader 8.0		
<b>Разработчик</b>	ScanSoft	ABBYY Software		
<b>URL</b>	www.pdffox.com	www.abbyy.ua		
<b>Стоимость</b>	от \$150	от 905 грн		
<b>Дополнительные программы</b>	PDF Converter, PDF Creator (в зависимости от версии)	ABBYY Screenshot Peaker		
<b>Системные требования</b>	Процессор – Pentium III и выше; ОС Microsoft Windows 2000/XP/Server 2003 Оперативная память – от 128 МБ Свободное место на жестком диске – от 270 МБ TWAIN-совместимый сканер, цифровой камера или факс-модем	Частота процессора – от 500 МГц ОС Microsoft Windows 2000/XP/Server 2003 Оперативная память – от 128 МБ Свободное место на жестком диске – от 350 МБ TWAIN-совместимый сканер, цифровой камера или факс-модем		
<b>Достоинства</b>	Богатая функциональность, возможность обработки PDF-документов с помощью дополнительных утилит	Удобный интерфейс с поддержкой русского и украинского языков, высокое качество распознавания, поддержка 179 языков, возможность обработки комбинированных документов с русскими и английскими словами, распознавание цифровых снимков и копий экрана		
<b>Недостатки</b>	Англоязычный интерфейс, малая доступность программы на украинском рынке, необходимость освоения системы, настройка	Последняя версия пакета не поддерживается операционными системами Windows 9x/Me		
<b>Скорость и качество распознавания</b>				
<b>Business Contact Manager</b>	<b>Время распознавания</b>	<b>Количество ошибок, %</b>	<b>Время распознавания</b>	<b>Количество ошибок, %</b>
2 страницы английского текста, заполнение текстом – 25%	56 с	5,29	1 мин 28 с	6,26
1 страница русского текста с изображениями, заполнение текстом – 40%	24 с	12,19	11 с	1,94
2 страницы русского текста, заполнение текстом – 70%	4 с	3,08	2 с	3,42
1 страница русского текста с изображением в верхней части, заполнение текстом – 80%	1 мин 47 с	28,32	21 с	7,70