

Как сохранить



В теперешний безбумажный, как любят его называть, век трудно представить, что какая-то информация еще доступна лишь на бумаге. Но наша страна в этом направлении делает только первые шаги, потому вопрос перехода на электронные носители информации останется актуальным еще довольно долго

Представьте себе центральную научную библиотеку с сотнями каталогов, огромным книгохранилищем и километровыми полками с подшивками периодики. Хотите такую же, но у себя дома в компактном виде — на ПК, ноутбуке, в электронной книге или КПК? Перевести нужную литературу в «цифру» своими силами вполне реально. Существует несколько простых устройств, методов и приемов, освоить которые не составит труда даже новичку. И нужная информация будет всегда под рукой!

Сканер-«ручник»

Возможно, первыми в голову и не придут портативные (ручные) сканеры, но такой класс устройств, пожалуй, может весьма пригодиться в «полевых» условиях.

Ручные сканеры так и не снискали популярности, скорее всего из-за своей специфики. Несколько лет назад их возможности были более чем скромные — большинство моделей могли лишь обрабатывать фотографии размером 10x15 см, а чтобы отсканировать целый лист формата А4 или больше, приходилось несколько раз сканировать его, последовательно обрабатывая участок за участком.

Примерно в то же время на рынке присутствовали и специализированные устройства для сканирования текста. Внешне они напоминали авторучку, текст обрабатывали построчно, что занимало слишком много времени, да и качество полученного результата оставляло желать лучшего. И те и другие сканеры требовали постоянного подключения к компьютеру, что, конечно, ограничивало фактор мобильности. Если требовалось, к примеру, пойти в библиотеку или архив, приходилось брать с собой ноутбук и ручной сканер.

Более новые устройства можно отнести к следующему поколению ручных сканеров — они сочетают в себе удобство и функциональность настольных решений, при этом работают независимо от компьютера. Рассмотрим, как же они работают, на примере одной из последних обновленных моделей подобных устройств — DocuPenRC800 от Planon System Solutions (<http://planon.com/>). В отличие от предшественницы DocuPenR700 новинка даже поддерживает цветное сканирование (может пригодиться при сканировании журналов или картинок).

В новых моделях ручных сканеров распознавание уже более точное, и пользователю не придется возить рукой по документу взад-вперед несколько раз. Достаточно 1–2 легких скольжений по бумаге



Подборку программ для чтения и конвертации электронных документов ищите на hi-Tech DVD

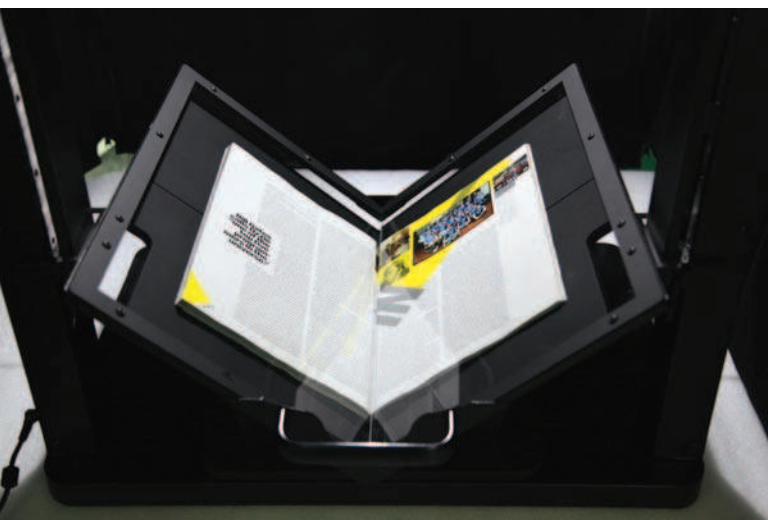
леса Амазонки

Встроенной памятью портативные сканеры обычно не блещут — в среднем до 8 МБ. Ее можно расширить за счет карт памяти microSD (до 2 ГБ — впрочем, для карманного устройства должно вполне хватать).

Внешне такой сканер выглядит как длинная палочка (или канцелярская ручка), примерно в два раза толще, чем обычная авторучка, а по длине приблизительно равная ширине листа бумаги формата А4. Переключение между режимами работы осуществляется функциональными кнопками, расположенными на корпусе устройств. Разрешение — от 100 до 400 точек на дюйм.

Конечно, вопрос удобства таких устройств довольно спорный. Но если сразу качественный результат при сканировании не получится, то после небольшого количества тренировок можно добиться надлежащей аккуратности.

Возможно, если бы на корпусе имелась специальная рукоятка или что-то подобное, пользоваться сканерами было бы еще удобнее. Но тогда эти устройства уже не будут столь компактными. А сейчас их спокойно можно брать с собой в библиотеку, в офис, в командировку. Правда, сегодня серьезную конкуренцию им могут составить компактные цифровые фотоаппараты, которые повсеместно вытесняют сканеры с рынка. Тем не менее, для ввода текстов пользоваться портативным сканером все-таки удобнее.



Так выглядит крепеж сканера, работающего с двумя фотокамерами вместо сканирующего элемента. Такой угол расположения книги не повреждает даже старинные фолианты

Chapter 9 Recycle This Code

This Chapter

- Using the AWT graphic user interface tools
- Responding to the user's interaction
- Working with GUI layout structures
- Reading Java Class Library documentation

The Java Class Library provides a number of prewritten classes that you can use directly to create useful applets. Among the most useful are the basic graphic user interface objects, such as buttons, check boxes, and drop-down lists. This chapter gives you a guided tour through these GUI objects and through the Java Class Library documentation that you will refer to when you use them.

Getting Started

- To set up a graphical user interface for an applet, you must take care of two issues:

- Putting the GUI objects on the screen
- Responding to user input

Chapter 5 gives you an example of how to display and respond to a button. The Java Class Library documentation tells you everything that you need to know to do it again.

Looking it up

To use a class that someone else has written for you, you need to know its constructors, its public methods, and something about what it does. For classes in the Java library, you can find all this necessary information in the

Примерно такое качество текста для распознавания мы получаем при работе с ручным сканером и после обработки прилагающимся к нему программным обеспечением, удаляющем неравномерные участки и различные «шумы»

О других ручных сканерах

Портативные сканеры C-Pen — www.cpen.com

Pegasus PC Notes Taker — www.pegatech.com/? CategoryID=218

7 правил фотографирования книг

Поскольку наши библиотеки сканерами пока похвастаться не могут, то отсканировать нужные страницы можно попробовать самостоятельно — своим фотоаппаратом. Чтобы получить более-менее подходящее для распознавания изображение, нужно следовать всего нескольким правилам:

1. Для сканирования подойдут фотоаппараты хотя бы от 3 Мпикс. Съемка с мобильного телефона в любом случае будет менее качественной, и программа может потом просто не распознать текст.
2. Пользуйтесь режимом макросъемки либо ручным фокусом. Стоит перепроверять, не оказался ли текст в расфокусе.
3. Если имеется возможность, снимайте текст при дневном свете возле окна либо при направленном освещении (например, настольной лампе).
4. Вспышка не поможет сделать текст более четким, наоборот, скорее засветит его, так как белые страницы ее преломляют.
5. Снимать лучше с упора — упритесь локтями о стол, подложите под руку стопку книг или, положив книгу на стул, используйте его в качестве опоры для рук.
6. Чем темнее в помещении, тем больше выдержка, тем заметнее будет смазана фотография при шевелении рук — увеличьте чувствительность (примерно до 400 ISO).
7. Если будет много нераспознанных мест, отредактируйте фото в каком-нибудь графическом редакторе: подберите контрастность и яркость, добавьте четкости и обрежьте лишнее.



Сканеры документации дают более качественную картинку, но, к сожалению, они предназначены только для отснятия отдельно взятых листов бумаги, и никак не подойдут для библиотечного пользования

Не только фото

Нельзя проигнорировать тот факт, что многие пользуются для сканирования в «походных» условиях обычным цифровым фотоаппаратом, не приобретая для этого отдельное устройство.

Кстати, некая предприимчивая компания Atiz еще два года назад выпустила весьма необычное устройство под названием BookDrive DIY (<http://diy.atiz.com>), предназначенное для сканирования книг. Собственно девайс состоит из специальной базы с подвижным креплением, на которое устанавливается обрабатываемая книга. Для разглаживания и придавливания страниц используется V-образная стеклянная панель. Цифровые изображения BookDrive DIY получает посредством закрепленной над книгой фотокамеры. Иными словами, устройство не сканирует, а просто фотографирует оригиналы.

Сканер BookDrive DIY, если, конечно, его можно так назвать, должен заинтересовать в первую очередь библиотеки, правительственные ведомства, издательские дома и научные организации, работающие с большими объемами текстовой информации и, в частности, книгами.

Новая разработка этого ряда — BookDrive Mini — существенно меньший, но с более высоким качеством изображения сканер. В работе с книгами он гораздо удобнее обычного сканера — исключены искривление страниц, повреждение корешка книги и низкая производительность. BookDrive Mini тоже использует так называемый V-образный прижим, который держит книгу в открытом положении с углом 120°.

Компактные помощники

Стоит вспомнить еще о парочке подвидов портативных сканеров. Например, протяжные, наподобие Genius ColorPage-SF600. Они сгодятся для оцифровки бланков и отдельных страниц, возможно, визиток и других деловых бумаг.

Да, подобные аппараты не блещут производительностью, но они предназначены именно для оцифровки документов. Протяжные сканеры более громоздкие, чем сканеры в виде ручки, хотя толщина и высота аппарата в среднем не превышают 5 см, а в длину они чуть больше, чем ширина листа А4. И при этом тот же ColorPage-SF600 является полноценным протяжным сканером с разрядностью цвета до 48 бит и разрешением до 600x600 ppi. Конечно, в реальной эксплуатации такие параметры будут чрезмерными в случае оцифровки монохромных текстовых документов, но вот для цветных деловых бумаг понадобятся и точность 600 ppi, и 24-битовый цвет.

Визитку — в «цифру»

Еще один довольно интересный подвид портативных сканеров — карманный сканер для визитных карточек. Примеры таких устройств — концепт-прототип B-Scanner (пока не запущен в серийное производство) и серия IRISCard Pro (см. hi-Tech PRO 1/07, с. 46). Они выполнены в несколько разных формфакторах и отличаются поддерживаемыми функциями.

Первый девайс имеет размер стандартной визитной карточки, сканирует каждое новое поступление и сохраняет скан в формате рисунка. Далее электронные версии визиток сортируются согласно желанию пользователя и могут быть по очереди просмотрены на OLED-дисплее, разрешения и четкости которого вполне хватает для того, чтобы рассмотреть все детали карточки.

Пожалуй, такое устройство может сослужить неплохую службу многим деловым людям. Разве что два нюанса вызывают некото-



Технологии сканирования и распознавания визиток развиваются довольно медленно, но уверенно. Все чаще они используются в мобильных устройствах, хотя они остаются оптимальными для создания баз данных контактов в крупных компаниях

рые сомнения: во-первых, не каждую современную визитку можно отобразить на дисплее; и, во-вторых, некоторые визитные карточки сегодня делаются двусторонними, а значит, их придется сканировать дважды, что уже не так удобно.

Сканеры визиток IRISCard направлены на работу с визитками после сканирования. Такие устройства сканируют бумажные визитные карточки, распознают содержащуюся в них информацию и автоматически создают электронные контакты для последующей передачи их в менеджер контактов, например MS Outlook, Lotus Notes, Address Book, Entourage и др., и экспортируют данные в Excel.

Для довольно небольшого девайса (формат А6) IRISCard Pro работает довольно быстро — скорость сканирования может достигать 600 визитных карточек в час, а разрешение для них остается достаточно высоким (около 400 точек/дюйм). Кроме визиток на нем можно сканировать цветные фотографии формата до 10x15 с разрешением от 300 до 600 точек/дюйм.

Адресная книга нового формата

Где удобнее всего хранить отсканированные визитки? Правильно — в смартфонах, коммуникаторах или КПК, которые всегда под рукой. Кстати, первые утилиты, способные прочитать данные с визиток, появились в смартфонах в 2006 году, когда мобильные камеры более-менее научились фотографировать в макрорежиме.

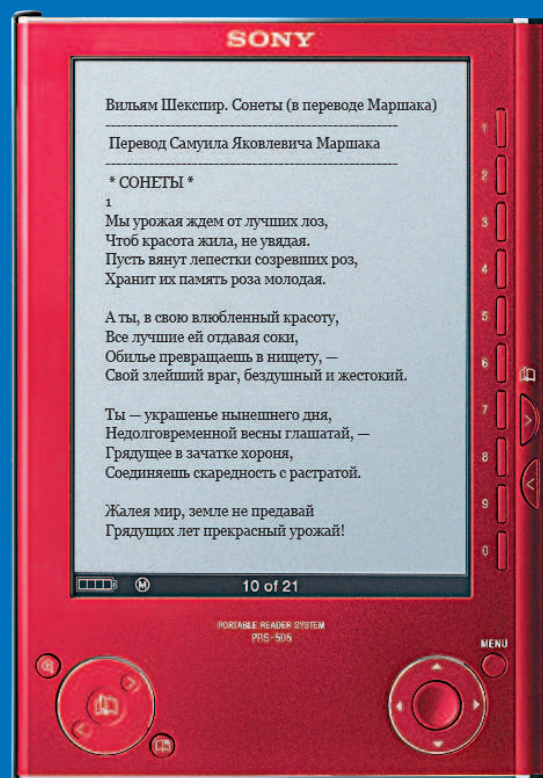
Одними из первых были коммуникаторы на базе операционной системы Windows Mobile — Samsung i600 и i710, но понимали они только англоязычный текст. Лишь Namecard Manager, созданный позже для коммуникатора E-Ten X500+, умел распознавать кириллические символы и вносить данные в соответствующие

Чем читать?

Мы уже писали о технологиях электронной бумаги (см. hi-Tech PRO 1/2009, с. 134 электронной версии). Для чтения отсканированного текста можно использовать практически любое устройство — от мобильного телефона до электронного ридера. Преобразовать текст в подходящий для девайса формат можно специальными конвертерами (http://reeed.ru/ru_converter.php). Кстати, у электронных книг Sony, например, есть свой собственный удобный формат — LRF (преобразовать в него проще всего с fb2 на сайте <http://onlinelrf.com>).

Нераспознанный текст можно сохранять в формате DjVu (от фр. *deja vu* — «уже виденное»). Так называют технологию сжатия изображений с потерями, разработанную специально для хранения сканированных документов — книг, журналов, рукописей и пр., где обилие формул, схем, рисунков и рукописных символов делает чрезвычайно трудоемким их полноценное распознавание. Также это будет эффективным решением, если необходимо передать все нюансы оформления — например, исторических документов, где важное значение имеет не только содержание, но и цвет и фактура бумаги; дефекты пергамента: трещинки, следы от складывания; исправления, кляксы, отпечатки пальцев; следы, оставленные другими предметами.

Для работы с этим форматом используются программы DjVu Reader, DjVu Editor или утилита WinDjView (<http://windjview.sourceforge.net/ru>).



Электронные ридеры, особенно с началом эры цветного e-ink, могут понемногу вытеснить КПК как устройство для чтения. Свидетельство тому — начало дизайнерских изысканий при создании e-book: от разноцветных до портативных моделей

Универсальный PDF

PDF остается наиболее универсальным для создания и хранения документов в электронном формате. С 1 июля 2008 года PDF даже стал открытым стандартом ISO (Международной организации по стандартизации, www.iso.org).

Главным преимуществом формата называют кроссплатформенность и интерактивность, ведь практически любой электронный текст на любой ОС можно преобразовывать и просматривать через любое устройство с установленным на нем Adobe Acrobat. В файлы PDF при желании можно добавлять комментарии, закладки, ссылки, аннотации, статьи, интерактивные элементы: кнопки и элементы формы (например, для проведения удаленного анкетирования), мультимедийные элементы.



Антон Правдивый,
 продакт-менеджер компании Samsung Electronics Украина

Мобильное распознавание визиток

В свое время компания Samsung показала очень удачную реализацию технологии распознавания визиток в мобильнике Samsung WiTu. Насколько эта технология оказалась востребована пользователями и как часто она применяется в других телефонах компании Samsung?

Пользователи отлично восприняли возможность распознавания визитных карточек, реализованную в ряде бизнес-моделей наших мобильных телефонов. И это неудивительно — данная функция действительно полезна, благодаря ей добавить новый контакт в телефонную книгу — дело даже не минут, а секунд. А временем деловые люди дорожат, как никто другой.



Хранители и распознаватели визиток в бизнес-моделях телефонов уже становятся не в диковинку, а скорее привычной функцией, которая наверняка не сегодня-завтра сможет переключаться и в бюджетные, и имиджевые модели

поля в записной книжке. Однако ошибались первые программы довольно часто, и некоторые данные приходилось вводить вручную, поэтому такие визитки не могли существенно экономить время.

Один из первых сканером визиток появился в коммуникаторе Sony Ericsson P990i. Приложение Card Scanner для Symbian UIQ 3.0 неплохо справлялось с распознаванием англоязычных карточек, «вытаскивая» оттуда имя и телефон владельца.

Смартфоны Nokia на Symbian S60 3rd Edition обзавелись встроенным сканером визитных карточек сравнительно недавно.



Михаил Бейрак,
руководитель отдела корпоративных проектов
компании «ABBYY Украина»

Когда придет электронный документооборот

Каким, по вашему мнению, будет следующий этап развития технологий распознавания текстов и изображений, способный значительно повысить удобство перехода с бумажной формы на электронную?

Развитие технологии распознавания текстов и изображений происходит в различных плоскостях в зависимости от целей преобразования:

— технологии распознавания изображений. Это, пожалуй, самая будоражащая воображение область. Сейчас практически используются технологии, позволяющие распознавать как похожесть предметов между собой, так и сходство лиц. Первые, например, используются таможенными службами ряда стран для быстрого определения ценности предметов, перевозимых за границу (достаточно сфотографировать объект и отправить его в центральную базу данных, в ответ будет найден предмет, внешне похожий, и указана его стоимость). В Украине аналогичные технологии используются для поиска похожих товарных знаков и изображений при проведении экспертизы интеллектуальной собственности. Поиск по лицам на изображении активно используется силовыми структурами ряда европейских стран;

— технологии полнотекстового распознавания. В этой области технологии развиваются по пути полного распознавания не только текста, но и оформления документа. Не только в плане

расположения слов и букв на странице, но и в плане правильного определения способа форматирования документа (автоматическое выделение верхних и нижних колонтитулов, автоматическое форматирование заголовков и построение их иерархии и т. д.). Также прогресс в этой области связан с разработкой распознавания все новых и новых языков;

— технологии классификации документов. В этой области важно правильно распознать тип документа, назначить соответствующий ему набор атрибутов и присвоить значения всем атрибутам. Здесь наблюдается тенденция реализации всех этих процессов без участия человека — к автоматическому определению типа документа и автоматическому извлечению необходимых атрибутов.

Эти технологии особенно близки к такой области использования распознавания, как извлечение данных из документов. Основной прогресс в этой области вращается как вокруг распознавания все новых и новых языков, штрихкодов, меток, так и вокруг технологий быстрой разработки алгоритма поиска объекта распознавания в документе. Собственно, именно в технологиях быстрой разработки алгоритмов поиска и ожидается наибольший скачок в ближайшие несколько лет.

Интерактивные PDF — удобное средство для сбора и анализа различных данных — в настоящее время используются еще недостаточно часто. Чего, по вашему мнению, не хватает этой технологии, чтобы она применялась более широко, в том числе на мобильных устройствах?

Это удобное средство, и мы поддерживаем разработку таких форм в своих продуктах. Однако в Украине их использование сдерживается порядком распространения электронно-цифровой подписи (ЭЦП). Дело в том, что любой документ, особенно электронный, становится интересным только при реализации алгоритмов, препятствующих подделке и искажению информации в нем. Именно таким атрибутом и является ЭЦП.

Утилита ABBYY Business Card Reader для распознавания данных пришла как нельзя кстати, правда, подходит она только для смартфонов, оснащенных камерой с разрешением не меньше 3 Мпикс. с автофокусировкой и макрорежимом.

Чтобы внести данные с визитной карточки в адресную книгу, надо сфотографировать визитку, проверить правильность распознанной информации и сохранить контакт в памяти телефона. Программа автоматически подставляет данные в нужные поля: фамилия, имя, должность и место работы, контактные телефоны и адреса. Процедура в среднем занимает 40–50 секунд. Правда, для Symbian OS утилита платная — 195 грн (<http://store.abbyy.ua/catalog/program.php?ID=52348>).

Эта же утилита доступна с прошлого года, например, для владельцев смартфонов Samsung WiTu (модель SGH-i900, работает на Windows Mobile 6.1 Professional). Кстати, в случае с этой моделью приложение достается пользователям бесплатно — оно встроено в телефон уже при покупке.

Документооборот 2.0

Идеология электронного документооборота сравнима с вебдванольной. Так, системы электронного документооборота (СЭД), по мнению экспертов, имеют ряд преимуществ:

- появление возможности коллективной работы над документами (что невозможно при бумажном делопроизводстве);
- значительное ускорение поиска и выборки документов (по различным атрибутам);
- повышение безопасности информации за счет того, что работа в СЭД с незарегистрированной рабочей станции невозможна, а каждому пользователю СЭД назначаются свои полномочия доступа к информации;
- повышение сохранности документов и удобства их хранения, так как они хранятся в электронном виде на сервере;
- улучшение контроля за этапами работы над документами.