

ОСОБЕННОСТИ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ РАСПОЗНАВАНИЯ

Знакомый незнакомец

Системы оптического распознавания символов иногда находят свое применение в самых неожиданных сферах. О том, что ждет нас в области распознавания, много говорилось на партнерском семинаре компании АBBYY Украина

Олег Пилипенко, pilipenko@chip.ua

Что такое OCR-система FineReader, знает, пожалуй, едва ли не каждый пользователь. Однако интеграция технологий распознавания в продукты сторонних разработчиков расширила сферу использования OCR-систем. Применение им порой может найтись в самых неожиданных областях. Впрочем, вы сами можете в этом убедиться.

Пишем, распознавая...

Иногда в документе необходимо распознать и конвертировать в «цифру» не весь текст, а только некоторые фрагменты, к примеру номер накладной или электронный адрес на визитке.

На помощь приходит ручка-сканер C-Pen, подключаемая к ПК или КПК через USB-порт. Надо лишь плавно провести ручкой по строке текста, и распознанный текст немедленно помещается в окне текстового редактора или в ячейку электронной таблицы (в общем, туда, где находится курсор). В C-Pen 600/800 используется ядро АBBYY FineReader, поэтому качество распознавания очень высокое. Поддерживаются 53 языка, 20 из них со словарем (включая русский и украинский).

MMS + OCR = сервис?

Устройства типа C-Pen требуют для подключения наличие USB-порта. Что же делать, если информацию надо передать, минуя компьютер, в мобильный телефон, где нет USB-интерфейса? С такой проблемой можно часто столкнуться при занесении информации с визиток в телефонную книгу мобилки. Решение уже найдено, хотя оно пока еще на стадии проекта. Необходимо сфотографи-

ровать визитку с помощью встроенной в телефон цифровой камеры и отправить в виде MMS на номер провайдера данной услуги. Через некоторое время вы получите распознанную информацию в формате телефонной книги именно вашей мобилки.

Для распознавания используется движок АBBYY FlexiCapture Studio. Эта высокоинтеллектуальная технология автоматически распознает, какое значение несет каждая строка на визитке (адрес ли это, телефон или же ФИО), и соответствующим образом обрабатывает все данные.

Фальсификациям — нет!

Наверняка в памяти многих украинцев еще свежи события прошлогодних выборов, на которых были обнаружены значительные фальсификации. А ведь технология распозна-



С помощью MMS и OCR-сервиса можно легко превратить информацию с визитной карточки в цифровой формат для телефонной книги

навания может поспособствовать в установлении достоверного результата. Методика, которую мы опишем далее, была использована на прошлых выборах в России.

Суть ее в том, что на некоторых избирательных участках в качестве эксперимента были установлены урны со



Ручка C-Pen 20, использующая ядро АBBYY FineReader, способна на лету распознавать фрагменты текста

встроенными сканером и устройством распознавания жестких форм на основе АBBYY FormReader (см. ЧИП 8/2003, с. 52). Когда избиратель бросал бюллетень, тот проходил через сканер и далее мгновенно распознавался. Каждые 15 минут OCR-устройство выдавало отчет о количестве проголосовавших, а также голосах «за» и «против». Конечно же, использование только этой технологии для подсчета голосов противоречит (пока что) украинскому законодательству. Тем не менее, применение такой методики для предварительного подсчета голосов в комплексе с надежной системой защиты данных позволяет снизить возможные ошибки и фальсификации практически до нуля.

Почитай мне вслух...

Средства для чтения электронных книг известны уже давно. А для людей с ослабленным зрением, благодаря использованию OCR-технологий, стало возможным создание устройств, читающих вслух бумажные документы. Машины «ИНФА» и «Книголюб-компакт» (а также западный аналог VERA System) содержат специальный сканирующий аппарат, голосовой синтезатор, а также систему распознавания на основе АBBYY FineReader. Аппарат умеет распознавать и анализировать страницы, озвучивать печатные документы, выполнять интеллектуальный разбор вслух таблиц и журнальных колонок с изображениями.

Мы описали только часть направлений, в которых могут использоваться и уже нашли применение технологии распознавания. Но даже этого вполне достаточно, чтобы увидеть, насколько широк круг применения OCR-систем и как они могут изменить нашу жизнь. 