

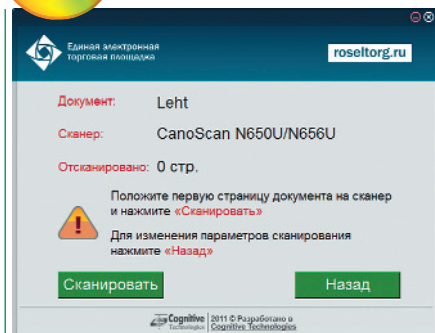


Стас Романов

УСКОРИТЕЛЬ БУМАГ



Обзор движка сканирования Cognitive ScanPack



Несколько секунд сканирования — и Cognitive ScanPack выдаст на-гора компактный PDF-файл

- Разработчик: Cognitive Technologies
- Веб-сайт: www.cognitiveforms.ru/products/scanpack

В нашу эпоху цифровых технологий обычный архив становится не только унылой архаикой, но и не способствует быстрому поиску нужной бумажки. А рабочее время, как известно, измеряется во вполне конкретных (и порой не менее бумажных) денежных знаках. Решение проблемы лежит на поверхности и сводится к переводу обычных документов в цифровую форму. Однако сканирование бумаги в большинстве случаев способно выдать на-гора либо распознанный текст (опционально — с сохранением таблиц, изображений и других элементов оформления), либо графические файлы.

К сожалению, технологии оптического распознавания (OCR) бумажных страниц часто пасуют перед документами со сложным оформлением, а в ряде случаев требуется оцифровка с полным сохранением исходного вида материала. Лишние манипуляции с дополнительной обработкой документов тоже не способствуют повышению производительности труда. «Бумага», сохраненная в графических файлах, пусть даже в сжатых форматах, занимает ощутимое место (экономия трафика при передаче данных никто не отменял) и не подлежит индексации для последующего поиска.

Оптимальным решением станет сканирование документов с последующим их сохранением в одном из кроссплатформенных форматов, например в PDF (формат DjVu так и не прижился в корпоративном секторе).

Отсканировать несколько страниц в домашних условиях — сущая мелочь. Временем, потраченным на эти манипуляции, можно пренебречь: все-таки сканер используется не каждый день. Другое дело — оцифровка бумажных документов в корпоративном секторе. Оборот документов в этом случае исчисляется сотнями, а то и тысячами страниц — тут без специального движка Cognitive ScanPack не обойтись.

Безусловно, для сохранения сканированного документа в PDF можно использовать OCR-приложения, но затраты времени на сканирование, распознавание и проверку текста и корректности форматирования далеки от желаемых.

И тут на помощь офисным труженикам придет программа Cognitive ScanPack, способная за считанные минуты не только отсканировать сотни страниц, но и улучшить качество полученных изображений и сохранить их в компактные файлы формата PDF. Более того, при использовании на предприятиях сторонних приложений возможно внедрение технологии Cognitive ScanPack с помощью программного интерфейса (API).

Данное решение будет весьма полезно при интеграции с ERP, СЭД, цифровыми архивами и системами, главной задачей которых является оцифровка документов при помощи сканера для дальнейшего сохранения оных в компактной форме. Типичный пример — солидные библиотеки, вознамерившиеся перевести свои издания в цифровую форму. Для оцифровки достаточно обычного сканера (мы тестировали продукт на устройстве 10-летней давности). Другая родственная сфера применения Cognitive ScanPack — наполнение электронных хранилищ и архивов путем сканирования большого объема обычных, бумажных документов. Компаниям, предпочитающим электронный оборот документов, тоже не обойтись без прогрессивной технологии, особенно в тех случаях, когда необходимо распределенное сканирование в филиалах для последующей передачи файлов в центр обработки.

Теперь о том, как это работает. Cognitive ScanPack умеет не только сканировать бумажные документы (в т.ч. возможно пакетное сканирование), но и работать с изображениями в форматах TIF, JPEG и PNG. Как и в любой другой программе сканирования, доступен быстрый просмотр изо-

бражений с опциями остановки и продолжения процесса. При необходимости можно задать свой вариант автоматического присвоения названий отсканированным файлам.

Разумеется, не все сканируемые бумажные документы могут похвастаться идеальным качеством печати (да и фактура листов подчас оставляет желать лучшего). К вашим услугам настройка профайлов автоматической обработки графических файлов с опциями поворота на 12, 90, 180 и 270 градусов; обрезки; устранения искажений и шумов; инвертирования, а также выбор между цветным изображением, черно-белым и оттенками серого цвета.

Выше мы говорили о компактном размере PDF-файлов, полученных после сканирования. Небольшой размер готового результата обусловлен фирменной технологией интеллектуального сжатия Cognitive PDF/A. Если быть точным, то готовый продукт сохраняется в файлах кроссплатформенного формата PDF/A (стандарт ISO 19005-1:2005), обеспечивающих не только безопасность при обмене данными, но и быстрый поиск в текстовом содержимом. Индексация для последующего поиска возможна за счет отдельного слоя текстового контента файлов PDF/A: дело в том, что в Cognitive PDF/A встроено ядро OCR-системы CuneiForm (www.cuneiform.ru).

Механизм оцифровки предельно прост: нужно загрузить в сканер бумажный документ, дать имя будущему PDF-файлу, указать папку для сохранения результата и нажать кнопку «Сканировать». Через несколько секунд вы получите предельно компактный PDF-файл. Воистину все гениальное просто. На нашем диске, к слову, вы найдете сокращенный, без временных ограничений релиз данной системы для Единой электронной торговой площадки (roseltorg.ru), который предназначен для работы с черно-белыми документами с подписями и печатями, содержащими не более двух цветов.