

## Клерк, ликуй!

### Российская компания обещает революцию в электронном документообороте

*Давно замечено, что переход на электронный документооборот приводит не к снижению, а к увеличению потребления бумаги, причем поистине катастрофическими темпами: в последнее десятилетие в России рост спроса на офисную бумагу составляет 12–18% в год, и пока не видно признаков его снижения. Архаичная российская законодательная база, ориентированная на бумажные технологии, тут, как это ни странно, ни при чем — на Западе в расчете на каждую «офисную душу» бумаги потребляется еще больше, чем в России.*

Одна из главных причин, сдерживающих развитие чисто электронного документооборота, — несовершенство технологий хранения, пересылки и работы с документами в электронной форме. Эти технологии, по сути, предлагают выбор между двумя крайностями: сохранить документ в виде изображения или преобразовать его в текст. В первом случае сохраняется аутентичность электронной и бумажной копий, но работать с таким документом окажется невозможно: прежде всего в нем нельзя осуществлять поиск по содержанию, а значит, и размещать в автоматизированных базах данных. В него нельзя внести правку, сгенерировать на его основе новый документ, получить автоматическое резюме, даже просто переслать его по электронной почте затруднительно из-за большого объема.

Вы все это сможете сделать, переведя документ в текстовую форму, но при этом утеряется его идентичность оригиналу: исчезнут подписи и печати, рукописные вставки, резолюции начальствующих, графические логотипы и шапки, форматирование, цветной фон и другие определяющие особенности его представления. Существующие технологии перевода документов в цифровую форму, разумеется, предлагают тот или иной компромисс между этими крайностями, но в большинстве случаев они требуют такого количества ручного труда по доводке, что проще и дешевле оказывается хранить документ в бумажном виде, пересылать обычной почтой и по старинке передавать его из рук в руки под роспись.

Технология ScanPack, представленная российской компанией Cognitive Technologies в середине апреля, предлагает путь для разрешения этого противоречия, относящегося к крупнейшим парадоксам современности. Ее специалисты придумали способ представления документов в электронном виде, позволяющий сохранить любые существенные особенности внешнего вида

оригинала, и при этом сделать его доступным для автоматического поиска и индексации. Технология позволяет в десятки раз уменьшить объем документа, сохранив все подробности, и, что немаловажно, уничтожить при этом все лишние детали внешнего вида: следы скрепок, степлера и дырокола, помятые уголки, случайные огрехи печати, посторонние брызги и грязь.

Все эти процессы максимально автоматизированы и потребуют, возможно, лишь контроля со стороны пользователя. Программа сама «поймет» разницу между грязными пятнами, печатным текстом, штампом организации, рукописной резолюцией и фоновым логотипом и применит к каждой из этих разновидностей свой отдельный алгоритм обработки.

По уверениям специалистов Cognitive, технология уже внедрена в одной из страховых компаний, для сохранения в электронном виде полисов, заполненных вручную на самокопирующихся бланках с цветным муаровым фоном — то, что обычно бесполезно предлагать программам распознавания. В эффектной демонстрации журналистам был продемонстрирован таможенный документ, сплошь покрытый разноцветными штампами, печатями и подписями, так, что прочесть его содержание было нелегко и глазами — и программа все-таки с ним справилась.

Технология может быть полезна далеко не только в бизнес-среде и для государственных ведомств: ее, несомненно, оценят в библиотечном деле — для перевода в электронный вид печатных изданий с сохранением особенностей их оформления. Очень существенный момент в этой инновации — пользователям не придется у себя внедрять какие-то особые технологии сканирования и отображения документов. Разумеется, обработка ScanPack осуществлялась на собственных технологиях Cognitive, но в качестве исходного материала вполне годятся и результаты работы других программ распознавания, включая наиболее популярный в нашей стране Fine Reader фирмы ABBYY. А конечный результат представляется в стандартном формате PDF, бесплатная программа для чтения которого под названием Adobe Reader входит в обычное оснащение любого офисного компьютера.

Если все рассказанное воплотится на практике, то значение этого достижения выходит далеко за рамки повышения комфортности существования «офисного планктона» и снижения потребления бумаги. За все годы цифровой революции нас много раз кормили несбыточными мечтами: обещали и автоматический машинный перевод, и удобное голосовое управление компьютерами, и безошибочное распознавание лиц террористов в толпе и еще многое другое. Фактически, однако, прогресс в области интеллектуальных функций компьютеров как минимум последний десяток лет стоит на месте — инновации не выходят за пределы косметических усовершенствований вроде автоматического поворота изображения на экране планшета, в зависимости от ориентации, или возможности пользоваться пальцем вместо мышки. Многие, наверное, помнят дурашливую скрепку-помощника из Microsoft Office 97, от которой все с раздражением

отказывались, — вот это и есть характерный уровень «прогресса» в этой области, и с тех пор принципиально ничего не изменилось.

Появление таких технологий, как ScanPack, — признак того, что дело наконец сдвинулось с мертвой точки. Ну, а действительно ли воплотится это в реальные достижения, мы имеем шанс оценить своими глазами: Cognitive Technologies, обычно ориентированная лишь на государственные органы и крупные корпоративные заказы, на этот раз обещает выпустить в том числе и массовый продукт для нас с вами, «простых пользователей».

Юрий Ревич

обозреватель «Новой»

21.04.2011