

«ОБЛАЧНАЯ» АВТОМАТИЗАЦИЯ ОБРАБОТКИ ДОКУМЕНТОВ

Е.Н. Затейчук (Компания Техкомпас)

Представлена реализации on-line сервиса по распознаванию документов, базирующаяся на использовании продукта ABBYY FlexiCapture. Сформулированы преимущества от использования on-line сервиса. Приведен пример реализованного проекта по внедрению on-line сервиса в транспортной компании.

Ключевые слова: распознавание, верификация, on-line сервис, информационная система, обработка документов.

В любой достаточно крупной компании, имеющей большое число клиентов, вопрос хранения и учета документов и информации, содержащейся в них, стоит достаточно остро. В связи с чем уже ни для кого не секрет, что информационная система для учета и хранения документов — это не роскошь, а жизненная необходимость. Особенно это важно при работе с финансовыми документами. Когда дело касается учета и выставления счетов, возможность найти нужную информацию, в кратчайшие сроки найти и просмотреть нужный документ становится особенно важной. Ведь задержка в несколько дней может повлечь за собой серьезные штрафные санкции.

На помощь приходит информационная система компании, в которой хранятся данные о документах. Это удобно: хранить данные, анализировать, обрабатывать, использовать в дальнейшем (например, вносить данные о заказах для автоматического формирования счетов), обращаться к скан-копиям документов для уточнения какой-то информации.

И здесь возникает вопрос: как обычно осуществляется этап ввода документов в систему учета?

В большинстве случаев одинаково: имеется гора однотипных документов, данные из которых нужно завести в БД компании, и некоторое число квалифицированных, высокооплачиваемых сотрудников, которые вручную вносят информацию из документов в эту самую БД, то есть выполняют работу оператора ввода информации.

При обработке счетов-фактур пик поступления документов бывает в последние дни месяца, когда, закрывая месячную отчетность, поставщики массово присылают документы. В эти дни документов счетов-фактур может быть в 10 раз больше, чем в остальные дни месяца. И появляется необходимость в срочном порядке внести и обработать эти данные.

Но даже опытные операторы при длительной работе совершают ошибки. От монотонной однотипной работы число ошибок возрастает. А ведь эти документы еще нужно будет отсканировать и добавить скан документа к нужной записи в той же системе учета.

С точки зрения компаний, сталкивающихся с подобного рода проблемами, было бы гораздо правильнее автоматизировать этот процесс. Идеальный вариант — это когда документы просто кладут в сканер, нажимают некую кнопку «обработать», и дальше пользователь вообще не интересуется тем, что происходит в голове у машины — она все сделает сама. Конечно, это фантастика, но очень близкая

к реальности. Потому что такие умные машины уже есть.

Компанией ABBYY разработана технология ABBYY FlexiCapture, которая позволяет автоматизировать обработку документов, снизить операционные расходы и повысить качество обработки документов.

Предположим, у заказчика имеется несколько разных типов документов, например, анкета, счет-фактура, заказ. Каждый из этих типов документов имеет много разновидностей. В разных документах может меняться расположение полей, ориентация самого документа, оформление. Общее для всех документов этого типа — перечень полей, которые обязательно должны быть на любом подобном документе.

Задача состоит в том, чтобы определить тип документа, извлечь необходимые данные из указанных полей, скан документа переименовать в определенном формате для обеспечения однозначной идентификации документа и его изображения. После чего извлеченные данные и скан с новым названием передать в информационную систему заказчика.

Так как данные из этих документов необходимо сохранить в БД компании, а к скан-копии документа будет необходимость вернуться, то вносить эти данные и сохранять эту скан-копию (а также формировать ссылку на документ) все равно придется. Вопрос — как сделать это с меньшими трудозатратами?

Можно рассмотреть вариант покупки лицензии ABBYY FlexiCapture, установки и настройки ПО, обучения сотрудников работе с программой. Такой подход может быть рентабельным только при довольно большом потоке документов, так как сама лицензия недешевая. Если даже поток документов достаточно большой, но не регулярный, несколько месяцев в году документов много, а потом наступает затишье, то покупка лицензии также может оказаться не самым выгодным решением.

Вариант, который может порадовать своим оптимальным решением — перенести саму программу обработки документов в «облако». Этот вариант подходит тем, у кого не такой большой поток документов, чтобы покупать лицензию, кроме того, нет желания занимать этой работой своего сотрудника, которого к тому же нужно обучать, а на случай его отпуска или больничного, искать кого-то ему на замену. Использование облачного ПО также подходит тем, у кого поток документов достаточно большой, но нерегулярный. При покупке лицензии нужно платить за год пользования лицензией и не факт, что это будет выгодно при количестве обрабатыва-

емых документов данной компании. А также этот вариант подходит для тех, у кого есть необходимость разово распознать большое число однотипных документов.

Использование облачного ПО предполагает работу с постраничной оплатой только за то число листов, которое было реально обработано.

Это решение уже хорошо зарекомендовало себя на практике. Выглядит это так — заказчик сканирует пакет документов и отправляет по электронной почте. Программа проверяет почтовый ящик раз в несколько минут, сканированные документы импортируются в программу обработки. Программа определяет тип документа, для определенного типа извлекаются нужные, описанные при настройке, поля. Если документ был хорошего качества и все необходимые поля были заполнены так, как нужно, то документ сразу экспортируется в папку, откуда дальше данные будут перенесены в информационную систему компании — заказчика.

Если же документ не очень хорошего качества, есть поля, заполненные от руки, впечатанные поверх готового документа, грязные, нечеткие или заполненные некорректно, то программа должна гарантировать по максимуму корректность вносимых данных. Для этого при настройке программного комплекса под конкретного заказчика настраиваются правила для проверки правильности внесения данных в поля. Например, если существует дата начала и конца некоторого действия, то настраиваем проверку, что дата конца не может быть раньше даты начала. Такие проверки помогают найти ошибку распознавания для тех полей, для которых есть строгие правила. Или, например, номер заказа должен состоять из семи цифр и первая из них не может быть нулем. Это правило формата. Все эти проверки помогают распознать данные как можно правильнее.

Если в ходе обработки документа появились ошибки правил или формата, или какие-то символы распознались программой недостаточно уверенно, появляется необходимость верифицировать документ, то есть проверить, правильно ли программа поняла то, что написано в документе. Для этого практичнее всего использовать специалистов-верификаторов компании, предоставляющей облачный сервис. Быстро и качественно они проверят правильность распознавания документов и отправят документ на экспорт.

Данные могут выгружаться в разных форматах в зависимости от того, как это будет удобно для дальнейшей интеграции с информационной системой заказчика. Так как разные компании работают с разными информационными системами, то сказать заранее, как будет настроен процесс интеграции, сложно. В некоторых случаях заказчики предпочитают настраивать процесс интеграции самостоятельно. В любом случае, если данные получены, всегда мож-

но найти способ загрузить их в ту систему, с которой работает компания.

Преимущества on-line сервиса распознавания документов:

- заказчик делает только первый шаг — кладет документы в сканер и запускает процесс сканирования. Сканер сам отсканирует и отправит документы в электронном виде по указанному адресу. А дальше on-line сервис обработает, верифицирует, отправит на экспорт, пришлет уведомление о том, что документы обработаны;
- отсутствие необходимости занимать сотрудников нудной механической работой;
- сокращение числа ошибок по сравнению с ручным вводом данных;
- увеличение скорости ввода документов в информационную систему;
- сокращение суммарных расходов на обработку.

В 2012 г. компания Техкомпас внедрила on-line сервис обработки документов в транспортной компании ЗАО «Руссо Транс», где требовалось наладить обработку документации по перевозке грузов. При этом также было необходимо минимизировать расходы на обработку документов и освободить сотрудников от ручного труда.

Работы начались с анализа разных видов документов, которые требовалось обработать. Далее были разработаны определения документов, настроены правила и проверки, свойства импорта документов и экспорта извлеченных данных. А также настроена интеграция с БД компании-заказчика.

В течение работы несколько раз появлялась необходимость скорректировать определения документов. Например, оказалось, что разновидностей некоторых документов так много, что сам заказчик не всегда знает все варианты документов, с которыми придется встречаться. Поэтому приходится уже в ходе работы дорабатывать определение документа. Но такие вопросы решаются быстро по ходу дела и не мешают рабочему процессу.

Внедрение заняло чуть больше месяца. Теперь отсканированные документы по электронной почте отправляются на специальный адрес; далее они импортируются в ABBYY FlexiCapture, верифицируются, импортируются в информационную систему «Руссо Транс».

Бумаги на перевозку часто приходят из разных городов нашей страны. Бывают документы с низким качеством печати или сканирования, слишком темные, грязные и т.д. Номер квитанции может быть вписан вручную или даже напечатан поверх другого текста. В этом случае для обеспечения высокого качества данных очень важна верификация. Верификация является частью on-line сервиса по автоматической обработке документов, позволяет ускорить процесс их обработки. А полностью автоматизированный процесс распознавания документов позволяет компании-заказчику сосредоточиться на профильной дея-

тельности, не заставляя сотрудников отвлекаться на обработку документов.

В результате использования on-line услуги по автоматизированной обработке документов в ЗАО «Руссо Транс» отмечают существенное уменьшение числа ошибок и многократное повышение производительности труда сотрудников.

Таким образом, on-line сервис по распознаванию документов будет полезен всем компаниям, которые имеют большой поток типовых документов и хотят освободить своих сотрудников от ручного ввода информации в компьютер, и одновременно сократить число ошибок, при этом минимизировав расход на обработку документов.

*Затейчук Елена Николаевна – руководитель направления компании Техкомпас.
Контактный телефон 8(905)774-62-04.*

IBM TIVOLI ENDPOINT MANAGER FOR SECURITY AND COMPLIANCE

Компания IBM

Компания IBM предоставляет инструмент Security and Compliance Analytics (SCA), являющийся Web-приложением платформы Tivoli® Endpoint Manager, предназначенной для оценки рисков и обеспечения безопасности ИТ-инфраструктуры. SCA включает библиотеки и инструменты технического контроля, которые позволяют непрерывно в автоматическом режиме обнаруживать и исправлять ошибки, поддерживать конфигурацию ИТ-системы в работоспособном состоянии.

Ключевые слова: реальное время, безопасность, ИТ-системы, облачные технологии, антивирусная защита.

Сегодня организациям и предприятиям для развития бизнеса необходимо внедрение новых ИТ-сервисов. При этом часто задача требует решения в рамках сильно ограниченного бюджета. Многие организации отдают предпочтение внедрению облачных моделей ИТ-сервисов, чтобы получить все преимущества от их использования, включая уменьшение затрат и перевод расходов из разряда капитальных в операционные, а также масштабируемость, гибкость и эффективное использование человеческих и технологических ресурсов.

Однако облака могут представлять потенциальные риски для безопасности и конфиденциальности бизнеса. Эти проблемы являются основной причиной замедления темпов внедрения облачных технологий в промышленности и крупных организациях. Защита высоко виртуализированных сред от целевых атак и угроз, обеспечение безопасной совместной работы пользователей и защиты данных (изоляция, совместное использование), нехватка прямого контроля параметров безопасности и конфиденциальности являются основными растущими проблемами безопасности облаков.

Пакет Tivoli Endpoint Manager for Security and Compliance от IBM призван гарантировать безопасность конечных точек всей ИТ-инфраструктуры организации. Он помогает защитить устройства и предоставить регулирующим органам доказательства того, что все стандарты безопасности соблюдаются. Это простое в управлении и развертывании решение обеспечивает высокий уровень защищенности в среде с большим числом разнообразных устройств: от серверов до настольных ПК. Данное решение поддерживает ноутбуки, работающие через Internet-соединение, и специализированные устройства, такие как кассовые терминалы (POS), банкоматы и киоски самообслуживания.

Это решение способствует уменьшению стоимости и сложности управления ИТ-средой, мало влияет на функционирование самих устройств, обеспечивает повышение производительности труда и более комфортные условия работы пользователей.

Функций безопасности

Функции управления исправлениями обеспечивают широкие возможности по распространению патчей для Microsoft Windows, UNIX, Linux и Mac OS, а также для Adobe, Mozilla, Apple, Java и других приложений среди распределенных устройств независимо от их местонахождения, типа соединения или состояния. Один сервер управления исправлениями способен поддерживать до 250 тыс. устройств, сокращая время установки исправлений без ущерба для функциональности устройств даже по сети с низкой пропускной способностью или территориально распределенной сети. Отчеты, составляемые в режиме РВ, содержат информацию о том, когда и кем были развернуты исправления, а также автоматическое подтверждение того, что исправления установлены. Таким образом, формируется полное решение для развертывания исправлений с обратной связью.

Функции управления параметрами безопасности представляют собой обширную библиотеку технических элементов управления, которые помогают соблюдать требования по безопасности, определяя и исправляя параметры настройки безопасности системы. Библиотеки поддерживают непрерывное соблюдение основных правил безопасности, обеспечивают сигнализацию, восстановление и подтверждение восстановления не соответствующих правилам безопасности устройств. Они также гарантируют проверку всех конечных точек ИТ-инфраструктуры в режиме РВ. Эта функция доставляет полезную информа-