

Насыщение метаданными



Артем МОИСЕЕНКО,
руководитель службы технологий
и архитектурных решений,
«Логика бизнеса» (ГК «АйТи»)

Исследование консалтинговой компании PricewaterhouseCoopers и Iron Mountain (www.pwc.es/es/publicaciones/tecnologia/assets/Seizing-The-Information-Advantage.pdf) показало, что у руководителей 75% компаний самой разной величины, местоположения и секторов экономики присутствует ощущение того, что они «извлекают из информационных активов максимум». Реально же успеха в этом деле добиваются лишь 4%. Примерно 43% опрошенных заявили, что их компания «извлекает из информации мало ощутимые преимущества», а 23% полагают, что никаких преимуществ нет вообще.

Три четверти организаций не обладают навыками и технологиями, которые позволяли бы эффективно использовать информацию для того, чтобы оторваться от конкурентов. В трех компаниях из четырех нет аналитика данных, и лишь в четверти организаций, где аналитик все-таки есть, используют его со знанием дела.

Информация в различных форматах поступает в компании отовсюду. Некоторые документы



Александра ЯКОВЛЕВА,
директор по корпоративным продажам,
АВВУУ Россия

оказываются мертвым грузом, даже если они переведены в электронный вид и положены в электронный архив. Найти необходимые данные в огромном объеме информации, которая представлена в виде отсканированных документов, почти невозможно – существующие средства поиска не работают с изображениями. Компаниям нужна система, способная на всех этапах работы со сканированным документом насыщать электронный образ дополнительными метаданными, позволяющими в дальнейшем анализировать документы и обрабатывать их не хуже собственно электронных документов, создаваемых сразу в электронном виде. И так, по порядку.

Ввод и классификация

На этапе ввода бумажный документ сначала сканируется, далее текст распознается и после этого анализируется. На основании полученных данных в системе документооборота создается карточка документа с автоматически

Объем данных, накапливаемых компаниями, растет очень быстро. За последние два года было сгенерировано около 90% всей существующей в организациях информации. По оценкам аналитиков Gartner и IDC, меньше чем через четыре года объем данных может составить порядка 44 зеттабайт (44 трлн гигабайт), а затем удваиваться ежегодно. При этом большая часть из них – около 80% общего объема корпоративного контента – неструктурированные.

заполненными атрибутами и вложением. В зависимости от тех или иных распознанных атрибутов мы имеем возможность автоматически выбирать маршрут движения документа, в том числе формировать в соответствии с заготовленными шаблонами перечень поручений для ответственного исполнителя и соисполнителей. Также появляется возможность осуществлять атрибутивный и полнотекстовый поиск по обработанным документам.

Или, например, ежедневно сотрудники канцелярии вручную разбирают электронные обращения, которые поступают от граждан через портал Госуслуги (www.gosuslugi.ru). Когда их количество превышает 1000 в день, это становится настоящей проблемой. Такая задача отнимает массу человеко-часов. Оптимизировать весь процесс можно за счет автоматизации ввода, а затем и классификации информации. Для этого существует технология, которая позволяет интеллектуально классифицировать и распределять поток входящих данных.

Как в электронных обращениях, так и в уже распознанных данных из бумажных документов классификатор автоматически определяет тематику и тональность обращений. Из текста извлекается важная информация (метаданные) для последующего создания регистрационной карточки. Затем СЭД автоматически отправляет документ по маршруту, который настроен для определенной тематики. Маршрут может быть задан специалистами вручную. Далее сотруднику канцелярии остается только проверить корректность заполнения карточки и отправить в исполнение.

Систематизация архивных документов

Сегодня многим организациям нужно систематизировать свои архивы документов. Каждый бизнес-процесс требует сортировки информации по различным признакам: структурному, хронологическому, функциональному, отраслевому, тематическому, предметно-вопросному, географическому, авторскому и другим, а также по содержанию определенного текста. При этом для многих документов имеются файлы с их электронными графическими образами или минимальным составом метаданных. Зачастую они лежат без дела, и найти среди них нужный документ достаточно трудно, не говоря уже о подборке по нужной тематике и поиске или анализе статистических данных. Ручное структурирование таких архивов – абсолютно неподъемный процесс. Он может растянуться на годы и потребует работы отдельно нанятых сотрудников, что выливается в серьезные материальные расходы для компании.

Тем не менее данные из таких хранилищ документов, приведенные в порядок, крайне полезны для организаций. Быстрый доступ к анализу, статистике и прогнозированию позволяет не только сэкономить за счет оптимизации процесса обработки документов, но и получить выгоду от возможности принятия бизнес-решений на качественно новом уровне.

В результате удается организовать атрибутивный поиск, т. е. по полям или реквизитам, а также по тексту. Кроме того, есть возможность применения инструментов BI (Business Intelligence) для извлечения знаний из неструктурированной информации, представленной в виде текста.

Только необходимая информация

Одна из не менее важных проблем бизнеса – так называемый информационный шум. В компании поступает огромное количество данных, и найти то, что необ-

и быстро найти информацию для проверки договоров, которые у них сейчас в работе. Аналогичные технологии может использовать техподдержка или приемная, чтобы наладить обработку обращений в автоматическом режиме.

Немаловажную часть в работе современной организации занимает аналитика. Интеллектуальные технологии позволяют провести полный и точный анализ, без упущения даже незначительных, на первый взгляд, деталей. Так, например, аналитическая служба может быстро находить информацию, извлекать смысл из отчетов, поступивших из филиалов/офисов компании. Это позволяет

Возможности использования интеллектуальных технологий для организации документооборота в компании позволяют в большом объеме информации найти самую необходимую для конкретного бизнес-процесса и принять верное решение.

ходимо, бывает крайне сложно. Поиск полезных данных отнимает время и замедляет производственные процессы, принятие решений и, как следствие, приводит к финансовым убыткам и утрате конкурентных преимуществ.

Чтобы сократить время обработки каждого документа, система автоматически предоставляет необходимую сопроводительную информацию: аналогичные прецеденты, нормативно-справочную информацию и т. д.

Возможности использования интеллектуальных технологий для организации документооборота в компании позволяют в большом объеме информации найти самую необходимую для конкретного бизнес-процесса и принять верное решение.

Сотрудники юридических отделов, например, могут легко

привести данные в структурированный вид и комплексно проанализировать произошедшие события.

В итоге

Компания «Логика бизнеса» на базе технологии ABBYY Compeno реализовала инструмент, который позволяет реализовать описанные выше возможности по работе с неструктурированной информацией в огромных массивах данных. С помощью разработанных специалистами компании технологий клиенты могут оптимизировать такие ресурсоемкие задачи, как ввод и классификация документов для СЭД и других корпоративных систем, перевод в цифровой вид и систематизация архивов документации, поиск необходимой информации в многообразии корпоративных данных для решения бизнес-задач. ■