

Леонов А. В.

al@tt.ru

Динамический документ – ключевой объект современных информационных систем

Аннотация

Цель данной статьи – показать, что ключевым объектом современных информационных систем, прежде всего систем управления документами, становятся динамические документы, то есть документы, создаваемые системой по запросу пользователя на основе доступной информации (в отличие от обычных электронных версий бумажных документов). Автор критически анализирует популярный термин "управление знаниями" и показывает, что качественным отличием модных в последнее время "систем управления знаниями" от традиционных "систем управления документами" является именно существенное расширение возможностей генерации динамических документов. В статье рассмотрено существующее положение дел в сфере электронного документооборота и дан прогноз магистральных направлений развития технологий работы с документами.

Введение

В последнее время на рынке систем управления документами наблюдается несколько тенденций. С одной стороны, программные продукты, которые изначально разрабатывались для автоматизации разных аспектов управления документами, постепенно сближаются друг с другом по функциональности, интегрируются с продуктами других производителей¹. С другой – в повседневный оборот входят такие термины, как "виртуальный" (virtual), "живой" (live, alive), "оперативно доступный по запросу" (on-line), "эволюционирующий" (evolving) или "интеллектуальный" (intelligent) документ, что находит отражение в отчетах исследовательских групп² и публикациях научных сообществ³. Наконец, все больше компаний начинают позиционировать свои продукты как "системы управления знаниями"⁴. Все эти процессы глубоко взаимосвязаны друг с другом и отражают разные стороны одного явления: в области управления документами происходит качественный переход, центральным моментом которого является возникновение нового объекта – динамических документов.

¹ Приобретения: Hummingbird [89] + PC DOCS + Fulcrum (1999), Documentum [87] + eRoom (2002), Vignette [113] + Epicentric (2002), Oracle [102] + PeopleSoft [104] + J. D. Edwards [95] (ожидается) и т. д. Интеграция: Documentum [87] + Verity [112], SAP [106] + Documentum [87], Convera [82] + Oracle [102] + SUN [110] и т. д.

² Gartner [88], IDC [93], Delphi Group [83], META Group [97], Ovum [103] и др.

³ IEEE Computer Society [57], IEEE Communication Society [58], IEEE Professional Communication Society [59], Association for Computing Machinery [54], American Society for Information Science and Technology [52], Association of Knowledgework [55] и др.

⁴ Например, IBM Lotus [92], Microsoft SharePoint Portal [98], Oracle Collaboration Suite [102], Sun Microsystems ONE Portal Server [110], Documentum [87], Hummingbird [89], Convera [82], Hyperwave [90], Open Text [101], Divine [85] и т. д., из отечественных - Cognitive Technologies ("Астарта") [81], Галактика ("Галактика-ZOOM") [70] и др.

До тех пор, пока речь шла об управлении электронными версиями бумажных документов, в центре внимания были задачи "автоматизации документооборота" или "автоматизации управления документами". Перевод оборота документов в электронную форму позволил многократно упростить процесс подготовки и согласования документов, ускорить их доставку адресатам и прохождение документами своего жизненного цикла, усовершенствовать систему хранения и поиска документов [6] – но не добавил ничего принципиально нового в сам процесс передачи информации от одних людей к другим посредством документов. Электронный документ создавался, утверждался, регистрировался, передавался на исполнение, двигался внутри организации и попадал в архив практически точно так же, как и бумажный: разница заключалась лишь в удобстве обращения с ним. Пользователь мог получить "из компьютера" только те документы, которые были кем-то когда-то "в компьютер" занесены – тексты, таблицы, изображения, аудио- или видеофайлы и т. д.

По мере развития технологий хранения, поиска и автоматической обработки информации появилась возможность не просто выдавать пользователю те документы, которые были когда-то кем-то занесены "в компьютер", а создавать по запросу пользователя новые документы, удовлетворяющие его потребности в той или иной информации [48]. Все чаще электронные версии бумажных документов хранятся в реляционных или объектно-ориентированных базах данных в виде отдельных информационных блоков, сведений об их характеристиках и взаимосвязях с другими блоками. Человек, желающий получить из электронных хранилищ данных интересующую его информацию в удобной форме (то есть в форме документа), более не ограничен набором тех документов, которые были занесены в эти хранилища: системы управления документами постепенно "научаются" автоматически генерировать на основе доступной информации новые документы с требуемыми характеристиками [21], [45].

Поэтому ключевым объектом современных информационных систем становятся динамические документы – "эволюционирующие", "интеллектуальные", "живые". Акцент сместился от "производителя" информации к "потребителю": основная проблема на данный момент – не автоматизировать оборот существующих документов (эта задача уже практически решена), а автоматически создавать документы, содержащие ту информацию, которая требуется конкретному пользователю. Практически все компании, предлагающие системы управления документами, так или иначе работают над этой задачей⁵; ее решение неизбежно повлечет глубокую перестройку существующих бизнес-процессов и схем работы с информацией.

Определение понятий

Поскольку в данной статье будут употребляться такие термины, как "информация", "данные", "знания", "документ", "управление знаниями", "управление информацией", "управление документами", автор считает необходимым в явном виде определить, что он будет понимать под этими понятиями.

Как известно, понятие "информации" не имеет общепризнанного определения. Автор будет опираться на одну из наиболее распространенных формулировок: "информация – это набор сведений (сигналов, символов), который уменьшает степень неопределенности у

⁵ В частности, IBM [92], Microsoft [98], Oracle [102], Documentum [87], Hummingbird [89], Convera [82], Open Text [101] и др., из отечественных - Cognitive Technologies [81], Оптима [73], ЛАНИТ [71], Digital Design [84] и др.

их получателя". Информацию могут содержать любые объекты или явления внешнего мира, воспринимаемые человеком через органы чувств: тексты, изображения (картины, скульптуры), звуки (речь, музыка), прикосновения, запахи, вкусы и т. д.

"Данные" часто определяют как "сведения, представленные в определенной знаковой системе и на определенном носителе". Однако, уловить отличия "данных" от "информации" при такой формулировке настолько сложно, что для практического применения это определение не годится. Автор будет понимать под данными информацию, представленную в электронном (цифровом) виде: например, оцифрованное изображение "Мона Лизы" или аудиотрек "Лунной сонаты" на компакт-диске.

"Знания" – философская и потому наиболее трудноуловимая категория. Расхожих определений множество, нет смысла здесь их все цитировать. Автор будет пользоваться понятием "знания" в следующем смысле: "знания – это информация, воспринятая человеком" или "знания – это то, что возникает в процессе восприятия человеком информации". Вопрос о том, являются ли знания непосредственно воспринятой информацией или же результатом восприятия, в данном случае не принципиален, автор хочет лишь подчеркнуть "личностный" аспект знаний: восприятие одной и той же информации (например, трудов Эйнштейна) может приносить разным людям (читателям) различные по объему и глубине знания. Знания – это то, что находится "в голове" у человека; передать другим людям человек может только информацию.

Для понятия "документ" автор предлагает следующее определение: "документ – это информация, зафиксированная на материальном носителе в удобной, естественной для человеческого восприятия форме". Таким образом, люди передают информацию друг другу либо лично (разговор, мимика, жесты и т. д.), либо посредством документов (текст, картина, скульптура, аудио- или видеозапись и т. д.). Не всякий объем информации, зафиксированный на материальном носителе, является документом. Молекула ДНК, безусловно, содержит огромное количество информации – но это не документ, она создана природой, и информация в ней записана оптимальным образом "с точки зрения природы". В то же время, расшифровка ДНК, записанная исследователем в форме, удобной для восприятия другими людьми (текст, таблицы, рисунки) – это документ.

Разумеется, понятие "удобства" достаточно субъективно: то, что удобно для одного человека, может быть крайне неудобно для другого (зашифрованное сообщение). Автор хочет подчеркнуть другое: документами следует считать все материальные объекты⁶, которые создают люди с целью зафиксировать информацию и которыми они обмениваются с целью ее передать, независимо от материала и формата. Юридическая практика также постепенно приходит к этому: в XIX веке под документом подразумевали только текст и графику (карты, схемы); в начале XX в понятие документа были включены фотография (паспорт) и кино ("документальные съемки"); сегодня мы свободно говорим о мультимедийных (многокомпонентных, составных) документах, которые содержат одновременно текст, графику, звук и видео. С началом массового использования объемных принтеров [114] в понятие документа неизбежно будет включена скульптура (рельефные изображения) и т. д.

В рамках данных определений рассмотрим, что следует понимать под понятиями "управление занятиями", "управление информацией" и "управление документами".

⁶ В том числе объемные распределения магнитных полей и последовательности электромагнитных импульсов, которыми на физическом уровне являются электронные документы.

Итак, знания – это то, что появляется "в голове" у человека в результате восприятия информации. Можно ли управлять самим процессом восприятия, менять его характеристики? Наверное, да, но это – задача крайне сложная, относящаяся к сфере психологии, воспитания, обучения и т. д. Гораздо проще управлять процессом получения информации человеком, определенным образом модулируя информационные потоки, которые поступают у нему. В этой сфере современная гуманитарная наука добилась заметных успехов (предвыборные технологии, управление общественным мнением и т. д.). Таким образом, на практике управление знаниями заключается в управлении потоками информации, поступающими к человеку.

Откуда человек может получать информацию? Если исключить из рассмотрения божественное откровение, "подключение к единому биоэнергоинформационному полю" и другие экзотические способы, остается три варианта: в результате непосредственного исследования (наблюдения) объектов окружающего мира, лично от других людей или из разного рода документов.

Непосредственное исследование человеком объектов окружающего мира поддается управлению хуже всего; фактически, единственный эффективный способ – это физическое ограничение доступа к тем или иным объектам. Личное общение человека с другими людьми поддается управлению несколько лучше; разработано немало методик по управлению коллективом, которые позволяют стимулировать возникновение ситуаций, когда люди могут лично обмениваться мнениями по определенным темам либо, наоборот, затруднить возникновение таких ситуаций. Однако наиболее впечатляющие результаты на данный момент достигнуты в сфере управления потоками документов, поступающих к человеку (включая СМИ), прежде всего благодаря развитию "высоких" технологий.

Резюмируя, автор хотел бы подчеркнуть: на данном этапе развития науки мы можем "управлять знаниями" только в той мере, в которой можем управлять информационными потоками, поступающими к человеку; в управлении же информационными потоками ключевую роль играет управление потоками документов. При этом из всей области "управления знаниями" только управление документами, поступающими к человеку, поддается автоматизации с использованием современных технологий.

"Безбумажный" документооборот и "управление знаниями"

Почему же тогда в последнее время все большее число производителей, ранее предлагавших средства автоматизации управления документами, начинают позиционировать свои продукты как "системы управления знаниями"⁷? Что стоит за этим – маркетинговый ход или реальные изменения в функциональности предлагаемых продуктов? Для того, чтобы ответить на этот вопрос, нужно вспомнить, как развивались системы автоматизированного документооборота и технологии управления документами.

Процесс перевода документооборота в электронную форму в ведущих западных странах начался несколько десятков лет назад. Для того, чтобы стал возможен масштабный переход от оборота "бумажных" документов к обороту их электронных аналогов, потребовалось развитие целого ряда технологий.

⁷ IBM Lotus [92], Microsoft SharePoint Portal [98], Oracle Collaboration Suite [102], Documentum [87], Hummingbird [89], Convera [82], Open Text [101] и др.

Сканирование документов. В любой достаточно крупной организации за время ее существования накапливается архив бумажных документов, который может включать десятки и сотни тысяч наименований. Поток входящих в организацию или создаваемых внутри организации бумажных документов также может составлять сотни наименований в день. Очевидно, что переход к "безбумажному" документообороту был бы невозможен без создания технологий сканирования, позволяющих быстро получать изображение документа в электронном виде (прежде всего, для его последующего распознавания). Современные промышленные сканеры обладают производительностью в 100-150 страниц в минуту, что можно считать практически достаточным для решения задач сканирования документов [86]. Однако, в данной области есть и не вполне удовлетворительно решенные задачи, в частности, сканирование переплетенных или сброшюрованных документов.

Автоматическое распознавание, регистрация и индексирование документов.

Использовать электронное изображение документа в некоторых случаях уже намного удобнее, чем бумажную версию – например, его можно практически мгновенно переслать по электронной почте любому адресату, подключенному к сети Интернет. Однако для того, чтобы получить действительно электронный аналог бумажного документа, который можно свободно редактировать в электронном виде, необходимо автоматически "распознать" электронное изображение, полученное в результате сканирования. Для неструктурированного документа в простейшем случае "распознать" означает извлечь текст с сохранением разметки и оформления в отдельный файл стандартного формата либо в отдельное поле базы данных. Для структурированных документов (например, бухгалтерских форм) необходимо, кроме того, автоматически идентифицировать тип документа (сопоставить с одним из заданных типов) и его структуру (сопоставить поля отсканированного изображения – "код", "дата", "получатель" и т. д. с соответствующими полями БД).

После того, как документ распознан, система может автоматически зарегистрировать его в электронном архиве (занести реквизиты документа в соответствующие поля электронной "регистрационной карточки" документа) и выполнить его полнотекстовую индексацию (построить список ключевых слов документа). Задача распознавания текстов в настоящее время практически решена⁸; успешно решается задача автоматизированного ввода структурированных документов (автоматическое распознавание полей отсканированного документа и заполнение соответствующих полей БД)⁹. Менее гладко обстоят дела с распознаванием реквизитов неструктурированных документов – автоматически решить, где у такого документа заголовок, где подпись автора, а где – дата выпуска, существующие системы могут только в достаточно простых случаях. Интенсивно развиваются технологии индексации текстов, однако и здесь есть свои трудности: индексация имен собственных, дат, денежных сумм, а также понятий, состоящих из нескольких слов.

Хранение документов. Масштабный переход к электронному документообороту был бы невозможен без развития технологий хранения данных, которые в настоящее время позволяют хранить огромные объемы информации на достаточно компактных и исключительно надежных носителях. Именно невероятное увеличение емкости и

⁸ Из популярных программных продуктов можно отметить ABYY FineReader [77], Cognitive Technologies CuneiForm [81], I.R.I.S. Readiris [94], ScanSoft OmniPage [108] и др.

⁹ Из популярных программных продуктов можно отметить ABYY FormReader [77], Cognitive Technologies CuneiForm [81], I.R.I.S. Formiris [94], ScanSoft OmniForm [108] и др.

снижение цен на носители информации за последние 10 лет (а также развитие алгоритмов и технологий сжатия данных) позволили к настоящему моменту практически свободно работать с оцифрованным звуком и видео, включая их в традиционные текстовые документы, что в итоге привело к повсеместному признанию мультимедийных (многокомпонентных, составных) документов. Создание электронных архивов документов также вряд ли было бы возможно без интенсивного развития систем управления базами данных¹⁰. Несмотря на то, что традиционные реляционные и объектно-ориентированные базы данных изначально не проектировались для хранения именно документов, их функциональность и гибкость в настоящее время достаточны для организации электронных архивов документов любого размера.

Редактирование документов. Очевидно, что переход к "безбумажным" технологиям был бы невозможен без разработки программного обеспечения, позволяющего создавать и редактировать электронные документы любому человеку со средним уровнем компьютерной грамотности. Прогресс в этой сфере затронул практически каждого пользователя, базовые функции ПО типа MS Word доступны 5-летнему ребенку (что, несомненно, является выдающимся достижением на пути к электронному документообороту). Технологии управления версиями документов [10], в той или иной степени реализованные практически в любом современном ПО для редактирования документов, позволяют одновременно работать над одним документом нескольким пользователям, вносить и согласовывать исправления, формировать финальный документ на основе последних версий его компонент, созданных различными пользователями и т. д.

Однако, в этой области есть свои проблемы, прежде всего – проблемы с совместимостью документов разных форматов, которые возникают из-за фактического отсутствия единого открытого стандарта электронных документов и "закрытости" популярных стандартов, таких как MS Word DOC [98] и Adobe PDF [78]. Развитие открытых стандартов, таких как XML, SGML (HTML), TeX и т. п., безусловно, приближает нас к решению этой проблемы, однако, это – существенно разные форматы, предназначенные для решения разных задач (XML – разметка структуры документа для обмена данными между различными приложениями [67], HTML – разметка для представления документа в браузере пользователя [56], TeX – разметка для профессиональной подготовки документа к печати [7], [66]), и тенденций к их объединению пока не наблюдается. Кроме того, в большинстве организаций документы в этих форматах до сих пор составляют лишь малую часть от всех электронных документов.

Подтверждение подлинности (авторства) и целостности документов. Для того, чтобы документ имел юридическую силу, необходимо, чтобы существовал механизм подтверждения его подлинности (авторства). Традиционным бумажным документам юридическую силу придают личная подпись (виза), печать, фирменный бланк с водяными знаками и т. д. Без разработки аналогичных механизмов для электронных документов их сфера применения была бы существенно ограничена; в частности, невозможно было бы перевести в электронную форму бухгалтерию или официальную переписку. Достижения в сфере криптографии, прежде всего развитие технологий шифрования с открытым ключом, позволили создать надежный механизм подтверждения подлинности (авторства) электронных документов – электронную цифровую подпись (ЭЦП) [2].

¹⁰ Таких, как IBM DB2 [92], Microsoft SQL Server [98], Oracle [102], Borland Software Corporation InterBase [80], Sybase SQL [111], MySQL [61], PostgreSQL [64], технология Sun Microsystems JDBC [110] и др.

Принцип "визирования" документа с использованием ЭЦП заключается в следующем. Пользователь получает в одном из центров сертификации (которых в России в последнее время становится все больше) два так называемых "ключа" - открытый и закрытый (две последовательности символов, взаимно однозначно соответствующие друг другу). Закрытый ключ, записанный на дискете, хранится у самого пользователя со всеми необходимыми мерами предосторожности (например, в сейфе, подобно печати); открытый ключ может быть доступен большому числу людей. Когда пользователю необходимо завизировать электронный документ, он запускает на своем ПК специальную программу, которая на основе содержания документа и закрытого ключа вычисляет ЭЦП документа (последовательность байтов) и дописывает ее в тот же файл. Электронный документ, завизированный ЭЦП, может быть свободно прочитан любым лицом (если только он дополнительно не зашифрован), однако в него невозможно внести никакие изменения без нарушения подлинности ЭЦП. Получатель документа с ЭЦП может легко проверить его подлинность (авторство) и целостность с использованием аналогичной программы и открытого ключа.

Широкому использованию ЭЦП мешают проблемы не технологического, а юридического и психологического характера. В России федеральный закон "Об электронной цифровой подписи" был принят лишь в январе 2002 г., и проработка правовых вопросов в этой сфере остается достаточно слабой [3]. Психологически процесс визирования электронного документа с использованием ЭЦП существенно отличается от собственноручной постановки подписи на бумажном документе, в частности, далеко не так очевиден принимаемый при этом объем обязательств. Решать споры типа "Моя бабушка случайно нажала какую-то кнопку, а теперь вы отберете у нее дом?" еще предстоит научиться.

Защита документов. В традиционном "бумажном" делопроизводстве для передачи секретных документов используется фельдгерская служба, курьеры и т. п. В электронном делопроизводстве секретные сведения приходится передавать, как правило, по незащищенным каналам связи, и защита документов осуществляется с использованием технологий шифрования с симметричным либо с открытым ключом [13].

При шифровании с симметричным ключом для кодирования и раскодирования документа используется одна и та же секретная последовательность байтов (которая должна быть известна и отправителю, и получателю). Шифрование с открытым ключом принципиально ничем не отличается от описанной выше технологии ЭЦП, но только для шифровки сообщения отправитель использует открытый ключ получателя, а получатель расшифровывает полученное сообщение с использованием своего закрытого ключа (таким образом, каждый, кто знает открытый ключ получателя, может отправить ему зашифрованный документ, который никто, кроме получателя, прочитать не сможет). Шифрование с симметричным ключом на 3 порядка быстрее, чем шифрование с открытым ключом, что может быть существенно при больших объемах документов, но при этом возникает отдельная задача – как передать получателю документа секретный ключ. В настоящее время часто поступают следующим образом: само сообщение шифруется симметричным ключом, а этот секретный ключ шифруется с использованием открытого ключа.

Задача придания документу юридической силы часто совмещена с задачей защиты содержащейся в документе информации от посторонних лиц. Теоретически, для шифрования и для визирования документа может использоваться одна пара ключей: отправитель шифрует сообщение открытым ключом получателя и визирует своим закрытым ключом, получатель проверяет подлинность и целостность документа с помощью открытого ключа отправителя и расшифровывает документ с помощью своего закрытого ключа, и наоборот (таким образом, отправляя документ, пользователь

использует свой закрытый ключ для постановки ЭЦП, а получая – для расшифровки документа). Так и поступали до тех пор, пока не выяснилось, что такая схема позволяет злоумышленнику с использованием специальных приемов достаточно просто вычислить закрытые ключи. В настоящее время для шифрования и визирования документов ЭЦП, как правило, применяют две разные пары ключей [22].

При соблюдении требуемых мер предосторожности (конфиденциальность закрытых ключей) никакие существующие методы и средства не позволяют взломать документы, зашифрованные с использованием современных технологий и завизированные ЭЦП.

Интернет и электронная почта. Наверное, не будет преувеличением сказать, что электронный документооборот был бы в принципе невозможен без широкого распространения быстрых и надежных каналов электронной связи. Собственно, именно масштабное развертывание сети Интернет, активное использование технологий Интернет для построения корпоративных сетей (Интранет) и повсеместное внедрение систем типа электронной почты [18], [31], [49] стали основными факторами, определившими переход от "бумажного" к электронному документообороту. Возможность передавать документы из одной точки планеты в другую за несколько секунд практически бесплатно явилась настолько привлекательной, что стремление политических и экономических элит к ее реализации "потянуло за собой" развитие целого комплекса технологий.

Современные системы передачи электронных сообщений¹¹ позволяют не только передать электронный документ от отправителя к получателю в любой узел всемирной сети, но и организовать рассылку документов по списку получателей, назначить документу маршрут движения и организовать контроль за его продвижением, обеспечить автоматический анализ обновлений определенных документов и пересылку обновленных версий (или уведомлений) на заданный адрес и т. д. Технологические проблемы в этой сфере связаны, пожалуй, лишь с увеличением пропускной способности магистральных линий связи и развитием технологий доступа в магистральные сети ("последняя миля"), что необходимо для свободной пересылки аудио- и видеофайлов. Задача передачи текстовых сообщений между пользователями сети Интернет в настоящее время может считаться полностью решенной.

Печать документов. Строго говоря, возможность печати электронных документов нельзя назвать необходимой составляющей электронного документооборота. Однако на практике процесс перехода от "бумажных" документов к электронным сопряжен с преодолением многих трудностей юридического и психологического характера, и без развития технологий вывода документов на печать и разработки доступных по цене принтеров для массового использования этот переход вряд ли мог бы состояться за столь короткое время. Сегодня принтер есть не только практически в каждой организации, но и у многих домашних пользователей, и любой электронный документ при необходимости может быть легко распечатан на бумаге. Качество печати у современных струйных принтеров (2400x1200 dpi, цвет 48 бит) [1] вплотную приблизилось к качеству традиционной офсетной полиграфии (обычные журналы и буклеты – 2540 dpi, газеты – 1200 dpi), чего вполне достаточно для решения любых практических задач документооборота.

Уже к середине 90-х гг. XX в. развитие технологий сканирования, автоматического распознавания, регистрации и индексирования документов, устройств хранения данных,

¹¹ Microsoft Exchange Server [98], IBM Lotus Notes/Domino [92], Novell GroupWise [100], Oracle Collaboration Suite [102] и др.

реляционных и объектно-ориентированных СУБД, средств редактирования и печати документов, технологий шифрования, механизмов подтверждения подлинности и целостности документов, повсеместное распространение Интернета и электронной почты позволило в широких масштабах осуществить переход к электронному документообороту [39]. Интеграция перечисленных технологий в единые программные комплексы привела к созданию многофункциональных корпоративных систем управления документами, автоматизирующих различные контуры документооборота [29].

С технической точки зрения, к настоящему моменту задача перехода к электронному документообороту практически решена в таких сферах деятельности, как бухгалтерский учет, складской учет (ERP, Enterprise Resources Planning), логистика (SCM, Supply Chain Management; SRM, Supplier Relationship Management), учет контактов с клиентами (CRM, Customer Relationship Management), управление кадрами и других хорошо формализованных областях¹². Интенсивно развиваются корпоративные системы для управления неструктурированными документами, в том числе содержащими аудио- и видео компоненты, обеспечивающие работу с полным электронным архивом документов предприятия¹³. Ряд ведущих поставщиков стремится позиционировать свои системы как средства для полной автоматизации всех контуров документооборота предприятия, однако на практике обычно используются несколько систем разных производителей, дополняющих друг друга (например, SAP [106] и Documentum [87]).

Развитие технологий позволило не только автоматизировать все традиционные этапы работы с документами (подготовку, редактирование, согласование, утверждение, регистрацию, доставку адресатам, прохождение документами своего жизненного цикла, контроль исполнения, хранение и поиск в архиве), но и реализовать в системах управления электронными документами совершенно новую функциональность, которая была в принципе недоступна при работе с "бумажными" документами.

Гиперссылки, аудио- и видео-компоненты. Наиболее очевидным принципиальным отличием электронных документов от бумажных является возможность установления связей между документами с использованием гиперссылок, а также вставки аудио- и видео-компонентов. Хотя понятие "гипертекста" появилось еще в середине XX века, широкое использование гиперссылок в электронных документах началось с возникновением сети WWW в Европейском центре ядерных исследований (CERN) в 1991 году и разработки языка разметки HTML (HyperText Markup Language) [56] как основного формата представления документов в сети WWW. Сегодня гиперссылки поддерживаются во всех популярных форматах электронных документов: MS Word DOC [98], Adobe PDF [78], TeX [66] и т. д. Гиперссылки освобождают читателя от необходимости самостоятельно искать в сети упомянутый в тексте документ – для того чтобы он начал загружаться в браузер пользователя, достаточно одного щелчка мышью на соответствующей гиперссылке. Возможность вставки в текстовый документ аудио- и

¹² Из огромного числа программных продуктов, представленных на рынке, можно отметить Oracle E-Business Suite [102], PeopleSoft [104], SAP [106], Siebel [109], i2 [91], J. D. Edwards [95], Sage Group [105], Scala [107], Baan [79], Microsoft Axapta [98], Microsoft Navision [98], из отечественных - 1С: Предприятие [68], Галактика [70], Парус [74], ДЕЛЮ [76] и др.

¹³ Из огромного числа программных продуктов, представленных на рынке, можно отметить IBM Lotus [92], Microsoft SharePoint Portal [98], Microsoft Exchange Server [98], Oracle Collaboration Suite [102], Documentum [87], Hummingbird [89], Convera [82], Open Text [101], Vignette [113], Divine [85], из отечественных - "Евфрат-Документооборот" (Cognitive Technologies) [81], ОПТиМА-WorkFlow (Оптима) [73], LanDocs (ЛАНИТ) [71], DocsVision (Digital Design) [84] и др.

видео-компонентов существенно обогащает спектр приложения электронных документов, в частности, значительно повышает эффективность обучающих программ или электронных учебных пособий.

Технологии шаблонизации. Электронные документы все чаще хранятся в реляционных или объектно-ориентированных базах данных таким образом, что отдельные структурные элементы (информационные блоки) документа записаны в отдельных полях базы данных; кроме того, в той же БД хранится информация об характеристиках каждого информационного блока и его взаимосвязях с другими блоками. Отдельно хранятся шаблоны вывода, определяющие структуру и оформление документов разных типов. Когда от пользователя приходит запрос на определенный документ, система автоматически выбирает нужный шаблон вывода и информацию из БД, генерирует документ в стандартном формате (например, HTML) и предоставляет его пользователю¹⁴. Такая форма хранения позволяет намного более гибко работать с содержанием документов, чем хранение в виде единого файла [8]. Например, пользователь может получить в виде отдельного документа список всех заголовков и аннотаций документов определенной тематики, или все информационные блоки, содержащие заданные ключевые слова. Кроме того, такая схема многократно упрощает изменение дизайна документов: для того, чтобы изменить оформление всех документов одного типа, достаточно изменить только один шаблон вывода.

Управление жизненным циклом и "интеллектуальные агенты". Все чаще документы "берут на себя" управление отдельными этапами своего жизненного цикла (document workflow), становясь таким образом полноправными участниками бизнес-процессов предприятия [35], [44], [47]. Например, созданный руководителем подразделения документ может автоматически быть направлен на исполнение определенному сотруднику с учетом его загруженности (по результатам анализа очереди документов на исполнение), при этом руководитель сможет получать уведомления о принятии документа на исполнение, ходе работы над документом и результатам выполнения задания. Все большее распространение получают специальные программы, автономные или встроенные в документ – "интеллектуальные агенты" (information agents, intelligent agents, knowledge agents, mobile agents) [28], [42], [51], которые автоматически анализируют содержание документов и происходящие с ними процессы и по результатам анализа предпринимают определенные действия. Например, такие программы могут выдавать уведомления об изменениях в документах, пересылать обновленные версии документов заинтересованным пользователям по заданному либо динамически формируемому списку, информировать ответственных лиц об устаревании содержащейся в документах информации и т. д.

Интеллектуальный поиск и анализ документов. Современные поисковые машины обеспечивают не только поиск документов по их реквизитам - номеру, заголовку, дате выпуска и т. д. (как в обычном бумажном архиве), но и поиск по результатам автоматического анализа всего содержания документа (полнотекстовая индексация). При этом поисковая машина не просто выдает пользователю ссылки на все документы, где встречается указанное слово или словосочетание, а анализирует найденные документы и сортирует их таким образом, чтобы наиболее адекватные запросу документы находились вверху списка [16]. Та же поисковая машина (или корпоративная система управления

¹⁴ Из распространенных систем шаблонизации можно отметить Perl Template Toolkit [62], PHP Smarty Template Engine [63], ASP Template [53], Macromedia ColdFusion MX [96], Sun Microsystems JavaServer Pages Technology [110] и др.

электронными документами) может автоматически вести полную статистику обращений к каждому документу и по запросу пользователя сортировать списки документов по их популярности и индексам цитируемости [17]. Пользователь, подключенный к сети Интернет, более не ограничен архивом документов своего предприятия при поиске интересующей его информации – ему доступен весь огромный объем информации, представленной в свободном доступе во всемирной сети.

Автоматическое реферирование. Огромным потенциалом обладают технологии автоматического реферирования (automatic abstracting, automatic summarization), развитие которых с переменным успехом идет уже несколько десятков лет [11], [33], [37]. Реферированием называется краткое изложение содержания исходного документа (оригинала) в новом документе (реферате). Один из методов реферирования заключается в том, что из исходного документа извлекаются наиболее значимые фрагменты, которые затем соединяются друг с другом – то есть, реферат представляет собой набор выдержек из исходного документа. Более сложные методы предполагают анализ исходного документа и изложение его содержания "своими словами". В этом случае для реферирования текстов применяются технологии синтаксического разбора, технологии концептуального (семантического) анализа с использованием семантических сетей, фреймов и т. д. [5], а также генераторы текстов на естественных языках.

Современные системы¹⁵ позволяют достаточно хорошо решать задачи реферирования отдельных текстов, при этом объем реферата составляет 5% – 30% от объема оригинала. Таким образом можно получить, например, 5-страничную "выжимку" из 100-страничного доклада. Однако для большей части насущных задач в этой сфере до сих пор нет удовлетворительных решений [32]: к таким задачам относятся, например, составление рефератов на основе нескольких источников, реферирование с большой степенью сжатия (1% и менее), реферирование аудио- и видео документов (audio abstracting, video abstracting). Существующие системы позволяют сделать "выжимку" из видеофильма, отобрав по определенным критериям наиболее характерные кадры или моменты [19], [34], но задачи типа "автоматически изложить содержание 2-часового видеофильма в 1-минутном ролике" пока обсуждаются лишь чисто теоретически.

Машинный перевод. Технологии автоматического перевода текстов на естественных языках долгое время развивались практически независимо от средств автоматизации документооборота [9], [24], и до сих пор системы машинного перевода поставляются преимущественно в виде отдельных программных продуктов¹⁶ [4]. Однако по мере развития систем управления электронными документами машинный перевод, как и автоматическое реферирование, постепенно занимает в них свое место как одно из основных средств работы с документами. Несмотря на то, что основным языком международного общения до сих пор считается английский, повсеместного отказа от национальных языков при публикации документов не наблюдается – наоборот, число документов на китайском, японском, малайском, испанском, немецком, французском, языках Индии и арабских стран постоянно растет. Для того чтобы отслеживать изменения даже в какой-то одной профессиональной области, может потребоваться мониторинг документов на 5-6 языках, что делает совершенно необходимым наличие средств автоматического перевода в корпоративных системах управления документами.

¹⁵ Например, TextAnalyst (Microsystems Ltd.) [99], "Аннотатор" (МедиаЛингва) [72] и т. д.

¹⁶ Например, ПРОМТ (компания ПРОМТ) [75], СОКРАТ (компания Арсеналь) [69] и т. д.

Совершенствование технологий интеллектуального поиска и анализа документов, СУБД и технологий шаблонизации, автоматического реферирования и машинного перевода, разбора и генерации текстов на естественных языках привело к качественному переходу в процессе развития систем управления документами: у пользователя появилась возможность получать из системы документы, которые никто никогда в систему не заносил¹⁷!

Этот факт в действительности являет собой настоящую революцию не только в документообороте, но и в управлении информацией вообще: способность системы автоматически находить требуемые сведения в разбросанных по всему миру хранилищах информации, агрегировать их и представлять в удобной для человека форме имеет далеко идущие последствия, которые уже в ближайшем будущем могут кардинально изменить стандартные представления об организации бизнес-процессов. Именно поэтому ведущие производители систем управления документами сегодня начинают позиционировать свои продукты как "системы управления знаниями" – фактически, под этим подразумевается способность системы автоматически создавать документы, которые нужны конкретному пользователю.

Динамический документ как точка роста

Анализ последних тенденций в сфере электронного документооборота показывает, что разработка методов и технологий генерации динамических документов становится магистральным направлением развития современных систем управления документами [28], [46]. Ситуация, когда система выдает на запрос пользователя несколько тысяч ссылок на "подходящие" документы, пусть даже и отсортированные определенным образом, более не может считаться удовлетворительной. В ответ на один запрос система должна выдавать один документ – цельный, удобный для восприятия текст с иллюстрациями, аудио- и видео-вставками, снабженный гиперссылками на другие документы с более подробной информацией.

Подобная функциональность могла показаться фантастической еще лет десять назад – однако в настоящее время развитие технологий позволило вплотную приблизиться к ее реализации [20], [38]. Усилия, прилагаемые крупнейшими производителями¹⁸ в этом направлении, объем задействованных ресурсов и масштаб спроса со стороны бизнес-сообщества позволяют ожидать, что удовлетворительные решения такого рода будут широко доступны уже через несколько лет. Можно ожидать, что разработка и совершенствование методов и технологий генерации динамических документов будет идти по ряду основных направлений.

Во-первых, очевидна необходимость дальнейшего развития технологий хранения информации и специализированных "документно-ориентированных" баз данных, их трансформация к потребностям динамического документооборота [41]. Необходима дальнейшая стандартизация моделей электронных документов, решение проблем

¹⁷ Простейший пример – так называемые "динамические" веб-сайты, где HTML-страницы формируются по запросам пользователей на основе информации из базы данных и шаблонов вывода. В ответ на тот же запрос, скажем, через месяц система выдаст уже другой документ, в котором могут быть обновлены некоторые компоненты, автоматически изменены, удалены или добавлены гиперссылки на связанные документы, содержание которых, в свою очередь, также может измениться.

¹⁸ IBM [92], Microsoft [98], Oracle [102], Documentum [87], Hummingbird [89], Convera [82], Open Text [101] и др.

совместимости документов разных форматов, перевод неструктурированных документов в структурированную форму [65]. Несмотря на то, что задача представления всей накопленной человечеством информации в структурированной форме ("всемирная база знаний") вряд ли может считаться разрешимой за конечное время, такая работа может и должна проводиться в отдельных, наиболее важных профессиональных областях [60].

Во-вторых, необходимо дальнейшее развитие технологий поиска информации, интеллектуального анализа и индексирования электронных документов [26]. Все существующие методы автоматического анализа текстов – синтаксического (структурного), семантического (концептуального), лингвистического (стилистического) – должны быть задействованы таким образом, чтобы образ документа в поисковой системе (список ключевых слов, "регистрационная карточка" или что-то еще) в максимальной степени соответствовал реальному содержанию документа [30], [40]. Сегодня поисковые машины легко могут быть обмануты с помощью достаточно простых приемов, что широко используется для так называемой "раскрутки" веб-сайтов. Необходимо также развитие технологий сравнительного анализа текстов для автоматического определения первоисточника информации, установления маршрутов распространения информации и анализа вносимых при этом изменений [43].

В-третьих, необходимо дальнейшее совершенствование методик построения и технологий разбора запросов, развитие естественно-языковых интерфейсов [50]. Сейчас процесс поиска информации в сети является итерационным: пользователь строит запрос, анализирует полученные результаты, самостоятельно переформулирует запрос и т. д. – до тех пор, пока очередная итерация не даст удовлетворительного для него результата. Такая схема более не может считаться эффективной: система должна автоматически анализировать точность, полноту и непротиворечивость запроса пользователя и предлагать ему варианты уточнения запроса, например, с использованием технологий кластеризации (clustering) найденных по запросу документов [27], [36], анализа наиболее часто встречающихся запросов, статистики популярности и индексов цитируемости документов [17] и т. д.

В-четвертых, необходимо дальнейшее развитие технологий шаблонизации, разработка программного обеспечения для интеллектуального построения шаблонов и генерации документов. В системе должно храниться не 50 шаблонов для вывода 50 типов документов, а 50 основных элементов шаблонов, на основе которых система сможет автоматически создавать тысячи форм вывода, каждая из которых будет в максимальной степени соответствовать запросу конкретного пользователя и доступному для удовлетворения этого запроса набору информации [14]. Учитывая постоянный рост числа структурированных документов, должны совершенствоваться технологии графического представления структурированной информации: построение таблиц и графиков на основе численной информации из БД, определение трендов (тенденций) и формирование предсказаний (прогнозов) с оценкой их достоверности [23].

В-пятых, качественное развитие должны испытать технологии автоматического реферирования и машинного перевода – соответствующее программное обеспечение должно стать основной составляющей интеллектуальных систем управления документами следующего поколения [12], [15], [25]. Любые достижения в сферах хранения, поиска и представления информации будут недостаточными для решения поставленной задачи ("один запрос – один документ"), если система не научится автоматически выбирать из найденных по запросу данных наиболее важные, критически значимые сведения, факты и утверждения и переводить их на язык запроса. В то же время, удовлетворительное решение задач автоматического реферирования и машинного перевода хотя бы только для

текстовых документов даст в руки пользователям невероятно мощный инструмент, возможности которого в настоящее время мы вряд ли можем полностью осознать.

Выводы

В заключение автор хотел бы отметить следующее. Необходимо понимать, что активная популяризация термина "управление знаниями" и продвижение на рынок "систем управления знаниями" – это, прежде всего, удачный маркетинговый ход производителей систем управления электронными документами. Критический анализ показывает, что основная функция предлагаемых программных продуктов осталась прежней: автоматизировать работу с огромными объемами структурированных и неструктурированных электронных документов.

В то же время, функциональность этих систем действительно испытывает колоссальное развитие. Если раньше основные усилия разработчиков были сконцентрированы на автоматизации ввода "бумажных" документов в систему, развитии технологий хранения и поиска документов в базах данных и совершенствовании средств коллективной работы с документами, то сейчас акцент сместился на развитие технологий интеллектуальной обработки и анализа содержания документов, совершенствование средств поиска требуемых сведений в хранилищах информации и их представления в удобной для пользователя форме.

Несомненно, что в настоящее время происходит качественный переход к системам управления документами следующего поколения, центральным элементом которых становятся динамические документы – документы, генерируемые системой по запросу пользователя. Возможно, именно в таком виде в ближайшие годы будет осуществлена давняя мечта изобретателей вычислительных машин – организовать диалог между человеком и компьютером в удобной, естественной для человека форме.

Автор благодарит за полезные советы и обсуждения Станислава Клименко, Виталия Обернихина и Владимира Лудинова.

Литература и ссылки

Книги и статьи

- [1]. Асмаков, С. Струйные принтеры – 2002//КомпьютерПресс, №12, 2002: <http://www.compress.ru/Article.asp?id=3994>
- [2]. Баричев С. Г., Серов Р. Е. Основы современной криптографии. – М.: Горячая линия - Телеком, 2002: <http://www.cryptography.ru/db/msg.html?mid=1162993>
- [3]. Волков, Л. Настал ли час электронной цифровой подписи?//Byte, №3, 2003, с. 68-72: <http://www.bytemag.ru/Article.asp?ID=1549>
- [4]. Воронкова, Ю.; Чеповский, А. PROMT и СОКРАТ: концепции и влечения//Компьютерра, №30, 2002: <http://www.computerra.ru/offline/2002/455/19261/>
- [5]. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем. – СПб: Питер, 2000. – 384 с.
- [6]. Клименко С. В. и др. Электронные документы в корпоративных сетях. – М.: Эко-Трендз, 1999. – 271 с.

- [7]. Кнут Д. Е. Все про TeX/Пер. с англ. М. В. Лисиной. – Протвино: АО RDTeX, 1993. – 592 с.
- [8]. Леонов А. В. и др. Публикация динамических документов рекламно-информационного характера на корпоративном сайте//Электронный журнал "Исследовано в России", 101, с. 1148-1185, 2003: <http://zhurnal.ape.relarn.ru/articles/2003/101.pdf>
- [9]. Марчук Ю. Н. Основы компьютерной лингвистики. – М., Сигнал, 1999.
- [10]. Плискин Е. Л. Управление версиями в системах коллективного создания документов//Сборник трудов Института системного анализа РАН "Развитие безбумажной технологии в организационных системах", 1999: <http://www.cognitive.ru/innovation/sbornic1/index.htm>
- [11]. Хан, У.; Мани, И. Системы автоматического реферирования: Пер. с англ.//"Открытые Системы", №12, 2000: <http://www.osp.ru/os/2000/12/067.htm>
- [12]. Чеповский, А. Неразрешимая проблема компьютерной лингвистики//Компьютерра, №30, 2002: <http://www.computerra.ru/offline/2002/455/19249/>
- [13]. Шнайер, Б. Прикладная криптография. Протоколы, алгоритмы, исходные тексты на языке Си. – М.: Триумф, 2002. – 816 с.
- [14]. Abe, Y.; Suzuki, J.; Tashiro, G.; & Yamamoto, Y. Persona: a Framework to provide Adaptive Presentation for Web Documents. - IPSJ Summer Programming Symposium: <http://www.yy.cs.keio.ac.jp/~suzuki/project/persona/index.html>
- [15]. Adams, K. C. The Web as Database: New Extraction Technologies and Content Management. – ONLINE, March 2001: http://www.onlinemag.net/OL2001/adams3_01.html
- [16]. Brin, S; Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. – Computer Science Department, Stanford University, 1998: <http://www-db.stanford.edu/%7Ebackrub/google.html>
- [17]. Chakrabarti, S.; van den Berg, M.; Dom, B. Focused crawling: a new approach to topic-specific Web resource discovery. - Computer Networks, Amsterdam, Netherlands, 1999: <http://citeseer.nj.nec.com/chakrabarti99focused.html>
- [18]. Coffman K. G., Odlyzko A. M. Growth of the Internet. – Optical Fiber Telecommunications IV B: Systems and Impairments, I. P. Kaminow and T. Li, eds. Academic Press, 2002, pp. 17-56: <http://www.dtc.umn.edu/~odlyzko/doc/networks.html>
- [19]. Cooper, M.; Foote, J. Automatic Music Summarization via Similarity Analysis. 2002: <http://citeseer.nj.nec.com/cooper02automatic.html>
- [20]. Costello, D. For Knowledge, Look Within. - Knowledge Management Magazine, September 2000: <http://www.destinationkm.com/articles/default.asp?ArticleID=563>
- [21]. Fikes, R.; Jenkins, J.; & Frank, G. JTP: A System Architecture and Component Library for Hybrid Reasoning. Knowledge Systems Laboratory, 2003: http://www-ksl.stanford.edu/KSL_Abstracts/KSL-03-01.html
- [22]. Gutmann, P. Encryption and Security Tutorial: <http://www.cs.auckland.ac.nz/~pgut001/tutorial/index.html>

- [23]. Harney, J. Predictive analytics: forecasting future trends from existing data. – KMWorld Magazine, January 2003, Volume 12 - Issue 1: <http://www.kmworld.com/publications/magazine/>
- [24]. Hutchins, W. J. Machine translation: past, present, future. – Chichester (UK): Ellis Horwood, 1986. – 382 pp: <http://ourworld.compuserve.com/homepages/WJHutchins/PPF-TOC.htm>
- [25]. Hutchins, W. J. Machine translation today and tomorrow. – In Computerlinguistik: was geht, was kommt? Festschrift für Winfried Lenders, hrsg. Gerd Willée, Bernhard Schröder, Hans-Christian Schmitz. – Sankt Augustin: Gardez! Verlag, 2002, pp.159-162: <http://ourworld.compuserve.com/homepages/WJHutchins/>
- [26]. Khoussainov, R.; Kushmerick, N. Optimising Performance of Competing Search Engines in Heterogeneous Web Environments. 2003: <http://citeseer.nj.nec.com/583911.html>
- [27]. Kitsuregawa, M.; Wang, Y. Link Based Clustering of Web Search Results. 2001: <http://citeseer.nj.nec.com/wang01link.html>
- [28]. Klusch, M.; Bergamaschi, S.; Edwards, P.; Petta, P. Intelligent Information Agents: The AgentLink Perspective. – Austrian Research Institute for Artificial Intelligence, Vienna, Austria, 2003: http://www.springer.de/cgi/svcat/search_book.pl?isbn=3-540-00759-8
- [29]. Koulopoulos, T. M.; Frappaolo, C. Electronic Document Management Systems: A Portable Consultant. – McGraw-Hill, Inc., New York, 1995.
- [30]. Lahtinen, T. Automatic indexing: an approach using an index term corpus and combining linguistic and statistical methods. 2000: <http://citeseer.nj.nec.com/lahtinen00automatic.html>
- [31]. Lyman, P.; Varian, H. R. How much information? Internet – summary: <http://www.sims.berkeley.edu/research/projects/how-much-info/internet.html>
- [32]. Mani, I. Automatic Summarization. – Amsterdam, The Netherlands, John Benjamins Publishing Co., 2001. – 285 p.
- [33]. Mitkov, R. Anaphora resolution. – Longman, 2002: <http://pers-www.wlv.ac.uk/~le1825/Longman/>
- [34]. Ngo, C.-W.; Pong, T.-C.; Zhang, H.-J. Recent Advances in Content Based Video Analysis. – International Journal of Image and Graphics, 2002: <http://citeseer.nj.nec.com/522088.html>
- [35]. Prior, C. Workflow and Process Management, 2003: <http://www.wfmc.org/standards/docs.htm>
- [36]. Rüter, S. M.; Zervas, G. The Curse of Dimensionality and Document Clustering. 1999: <http://www.doc.ic.ac.uk/~srueger/pr-g.zervas-1999/abstract.html>
- [37]. Saggion, H. Automatic Abstracting: towards a Text Based Generation: <http://citeseer.nj.nec.com/465805.html>
- [38]. Sherman, L. Reinventing Document Management for Modern Business. - CRM Magazine, Tuesday, October 23, 2001: <http://www.destinationcrm.com/articles/default.asp?ArticleID=1859>
- [39]. Sutton, M. J. D. Document Management for the Enterprise: Principles, Techniques, and Applications. – Wiley Computer Publishing, New York, 1996.

- [40]. Williamson, B.; Miller, L. The semantic web: A touch of intelligence for the internet? 2003: <http://education.guardian.co.uk/elearning/story/0,10577,981948,00.html>
- [41]. Woods, E. Knowledge management 2002-2003: the end of the beginning. – KMWorld Magazine, January 2003, Volume 12-Issue 1: <http://www.kmworld.com/publications/magazine/>
- [42]. Wooldridge, M.; Jennings, N. Intelligent Agents: Theory and Practice. – Knowledge Engineering Review, Volume 10, No 2, June 1995: <http://www.csc.liv.ac.uk/~mjw/pubs/ker95/ker95-html.html>
- [43]. Zha, H.; Ji, X. Poster session: Correlating multilingual documents via bipartite graph modeling. – Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, 2002: <http://portal.acm.org/>
- [44]. AIIM User Guide. Business Process Management and Workflow: http://www.aiim.org/inform/all_docrep.asp
- [45]. AIIM User Guide. Document and Content Capture: http://www.aiim.org/inform/all_docrep.asp
- [46]. Описание семейства технологий Autonomy: <http://www.autonomy.com/Content/Technology/>
- [47]. CSC Index Research and Advisory Services. Foundation Report 112: Workflow and Business Process Design Tools: <http://www.cscresearchservices.com/foundation/library/112/RP01.asp>
- [48]. Delphi White Paper. The Document is the Process: <http://www.delphigroup.com/pubs/whitepapers/>
- [49]. The UCLA Internet Report. Surveying the Digital Future. Year Three. – UCLA Center for Communication Policy. 2003: <http://www.ccp.ucla.edu/pages/internet-report.asp>

Тематические веб-сайты

- [50]. Международный семинар Диалог: <http://www.dialog-21.ru/>
- [51]. American Association for Artificial Intelligence: <http://www.aaai.org/>
- [52]. American Society for Information Science and Technology: <http://www.asis.org/>
- [53]. Официальный сайт ASP: <http://www.asp.net/>
- [54]. Association for Computing Machinery: <http://portal.acm.org/>
- [55]. Association of Knowledgework: <http://www.kwork.org/>
- [56]. Официальный сайт HTML: <http://www.w3.org/MarkUp/>
- [57]. IEEE Computer Society: <http://computer.org/>
- [58]. IEEE Communication Society: <http://comsoc.org/>
- [59]. IEEE Professional Communication Society: <http://ieeepcs.org/>

- [60]. Knowledge Markup Language (KML) home page: <http://kml.mipt.ru/>
- [61]. Официальный сайт разработчиков MySQL: <http://www.mysql.org/>
- [62]. Perl Template Toolkit Home Page: <http://template-toolkit.org/>
- [63]. PHP Smarty Template Engine: <http://smarty.php.net/>
- [64]. Официальный сайт разработчиков PostgreSQL: <http://www.postgresql.org/>
- [65]. Resource Description Framework (RDF): <http://www.w3.org/RDF/Overview.html>
- [66]. TeX Users Group (TUG) home page: <http://www.tug.org/>
- [67]. Официальный сайт XML: <http://www.xml.org/>

Веб-сайты компаний

- [68]. 1С: <http://www.1c.ru/>
- [69]. Арсеналь: <http://www.ars.ru/>
- [70]. Галактика: <http://www.galaktika.ru/>
- [71]. ЛАНИТ: <http://www.lanit.ru/>
- [72]. МедиаЛингва: <http://www.medialingua.ru/>
- [73]. Оптима: <http://www.optima.ru/>
- [74]. Парус: <http://www.parus.ru/>
- [75]. ПРОМТ: <http://www.promt.ru/>
- [76]. ЭОС: Электронные офисные системы: <http://eos.ru/eos/>
- [77]. АБВУУ: <http://www.abbyy.ru/>
- [78]. Adobe: <http://www.adobe.com/>
- [79]. Ваан: <http://www.baan.com/>
- [80]. Borland Software Corporation: <http://www.borland.com/>
- [81]. Cognitive Technologies: <http://www.cognitive.ru/>
- [82]. Convera: <http://www.convera.com/>
- [83]. Delphi Group: <http://www.delphigroup.com/>
- [84]. Digital Design: <http://www.digdes.ru/>
- [85]. Divine: <http://www.divine.com/>
- [86]. Document Scanners: <http://www.highspeedscanner.com/>

- [87]. Documentum: <http://www.documentum.ru/>
- [88]. Gartner: <http://www.gartner.com/>
- [89]. Hummingbird: <http://www.hummingbird.ru/>
- [90]. Hyperwave: <http://www.hyperwave.com/>
- [91]. i2 Technologies: <http://www.i2.com/>
- [92]. IBM: <http://www.ibm.com/>
- [93]. IDC: <http://www.idc.com/>
- [94]. I.R.I.S.: <http://www.irislink.com/>
- [95]. J. D. Edwards: <http://www.jdedwards.com/>
- [96]. Macromedia: <http://www.macromedia.com/>
- [97]. META Group: <http://www.metagroup.com/>
- [98]. Microsoft: <http://www.microsoft.com/>
- [99]. Microsystems: <http://www.analyst.ru/>
- [100]. Novell: <http://www.novell.com/>
- [101]. Open Text Corporation: <http://www.opentext.com/>
- [102]. Oracle: <http://www.oracle.com/>
- [103]. Ovum: <http://www.ovum.com/>
- [104]. PeopleSoft: <http://www.peoplesoft.com/>
- [105]. Sage Group: <http://www.sage.com/>
- [106]. SAP: <http://www.sap.com/>
- [107]. Scala: <http://www.scala.net/>
- [108]. ScanSoft: <http://www.scansoft.com/>
- [109]. Siebel Systems: <http://www.siebel.com/>
- [110]. Sun Microsystems: <http://www.sun.com/>
- [111]. Sybase: <http://www.sybase.org/>
- [112]. Verity: <http://www.verity.com/>
- [113]. Vignette: <http://www.vignette.com/>
- [114]. Z Corporation: <http://www.zcorp.com/>