

Институт вычислительных технологий СО РАН
пр. Акад. Лаврентьева, 6, Новосибирск, 630090, Россия

Новосибирский государственный университет
ул. Пирогова, 2, Новосибирск, 630090, Россия
E-mail: fedotov@nsu.ru; fedotov@sbras.ru

ПАРАДОКСЫ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ *

Статья посвящена современным проблемам развития информационных технологий и эксплуатации информационных ресурсов в условиях сетевой глобализации общества.

Ключевые слова: информационные технологии, информационные ресурсы, история, сети, информационные революции, парадигмы программирования, поиск.

Введение

Проблема доступа к информации является одной из основных проблем, возникающих в современной человеческой деятельности. Любой производственный или научный процесс порождает огромные объемы данных, и работать с ними становится все сложнее по мере того, как гигабайты данных превращаются в терабайты. Количество данных когда-нибудь превысит способность компьютеров их обрабатывать, поэтому необходимы новые инструментальные средства и алгоритмы для их анализа. Объемы сохраняемых сегодня данных превысили все цифры, о которых говорилось в самых смелых прогнозах, и процесс их накопления будет только ускоряться. Вместе с тем предъявляются серьезные требования к обеспечению прозрачного доступа и долговременной сохранности «информации». А в результате вопросы «что хранить?», «как хранить?» и «как найти?» остаются самыми существенными: без ответа на них все остальные теряют актуальность.

Может сложиться впечатление, что развитие информационных технологий уже само по себе способно вывести работу с научной информацией на качественно новый уровень, но, к сожалению, это совсем не так. Современные информационные технологии пока не могут предоставить адекватный аппарат для оперирования с «информацией» и информационными ресурсами [1].

Нынешнюю технологическую революцию характеризует не центральная роль знаний и технологий, а применение знаний и информации к генерированию знаний и созданием систем, обрабатывающих информацию и осуществляющих передачу «информации».

Парадоксы информационных технологий

Появление сетей в человеческом обществе является жизненной потребностью общества обмениваться информацией. Начиная с далекой древности человечество стремилось найти средства для быстрой и надежной передачи сообщений ¹. Поэтому появление сетей является не столько научным достижением, сколько исторической закономерностью.

* Работа написана по материалам лекции, прочитанной автором на Международной летней школе молодых ученых государств-участников Содружества независимых государств «Интеграция и инновации в воспроизводстве кадров для развития гуманитарного сотрудничества стран СНГ» (20–24 июля 2008 г., Новосибирск).

¹ *Информация* происходит от латинского слова «*informatio*», что означает сообщение, сведения, разъяснения, изложение. Этот термин получил широкое распространение, однако понятие информации остается одним из самых дискуссионных. Несмотря на попытки найти общие свойства и закономерности, присущие многогранному

Возникновение сетей. Если не вспоминать сигнальные костры, гонцов или почтовые тройки, то первая надежная крупномасштабная сеть для передачи сообщений со стандартизированной системой кодирования появилась во Франции. Французским механиком Клодом Шаппом был изобретен оптический, или семафорный, телеграф. Передача информации происходила с помощью вращения перекладки вокруг своей оси, прикрепленной к металлическому шесту на крыше башни. Первая его линия была построена между Парижем и Лиллем в 1794 г. Русский механик-самоучка Иван Кулибин изобрел систему семафорного телеграфа, которую он назвал «дальновещающей машиной», с оригинальным сигнальным алфавитом и слоговым кодом. Изобретение Кулибина было забыто царским правительством. В России телеграф Шаппа с некоторыми усовершенствованиями был введен в 1833 г. и соединял Петербург с Варшавой. Эта линия работала довольно успешно. Для передачи небольшой депеши на расстояние 1 тыс. км требовался один час. Активно использовался оптический телеграф и во время Крымской войны.

Решающее слово в создании быстродействующего средства связи принадлежит замечательному русскому ученому П. Л. Шиллингу. В 1828 г. им был испытан прообраз будущего электромагнитного телеграфа. Шиллинг был первым, кто практически решил проблему создания кабельных изделий для подземной прокладки, способных передавать электрический ток на расстояние. Дело рано умершего Шиллинга продолжил академик Б. С. Якоби, который изобрел телеграфный ключ, телеграфный код и печатающее устройство. В истории электротелеграфа самым популярным стал американский живописец Сэмюэл Морзе. Он построил аналоговый² телеграфный аппарат и создал азбуку к нему, позволяющие с помощью нажатия на ключ Якоби передавать информацию на дальние расстояния. Благодаря простоте и компактности устройства, удобству работы телеграф Морзе в течение полувека был наиболее распространенной системой телеграфа, применявшейся во многих странах.

Развитием телеграфа Якоби стал стартовый аппарат Жана-Мориса Эмиля Бодо (1874 г.) и пятибитный цифровой код телеграфии, который стал прообразом современной системы кодирования информации в компьютерах (код ASCII). Фактически сеть международного абонентского телеграфа (телетайп) является первой глобальной сетью со своими правилами передачи информации (принцип коммутации каналов, который в начале использовался и в компьютерных сетях адресации и маршрутизации, и со своими стандартами кодирования информации, которые оказали сильное влияние на стандарты кодирования информации в компьютерных сетях).

Изобретения телеграфа, телефона и радио произвели первую «информационную» революцию в обществе: телефон оказался первой сугубо личной коммуникационной технологией, радиосвязь кардинально расширила понятия коммуникационной сети, а телеграф явился предшественником глобального информационного пространства.

Информационные революции. Появление компьютеров привело к самому большому скачку в развитии человеческого познания и, как следствие, к колоссальному увеличению объемов перерабатываемой информации. Если в 1800-е гг. сумма человеческих знаний удваивалась каждые 50 лет, в середине XX в. – каждые 10 лет, то к 1970 г. период удвоения сократился до 2–3 лет.

Этот экспоненциальный рост с легкой руки Альберта Эйнштейна называют «информационным взрывом» с тем отличием, что понятие «взрыв» означает самоликвидирующийся процесс, а развитие познания бесконечно. Причина его возникновения связана не столько с гигантским прогрессом в информационных технологиях, сколько с возросшим во много раз потоком информации, необходимой для жизни современного общества.

Развитие вычислительной техники не шло гладко и размеренно. С одной стороны, уже при построении ЭВМ первого поколения возник вопрос: может ли машина мыслить. С дру-

понятию «информация», оно во многом остается интуитивным и получает различные смысловые наполнения в различных отраслях человеческой деятельности. Наиболее бесспорным выглядит определение, данное А. Н. Комогооровым: «Информация есть характеристика не сообщения, а соотношения между сообщением и его потребителем. Без наличия потребителя, хотя бы потенциального, говорить об информации бессмысленно» [2].

² Телеграфный аппарат Шиллинга был цифровым.

гой стороны, внедрение ЭВМ встречало бурное сопротивление со стороны ее потенциальных пользователей.

Хронологически средства вычислительной техники возникли прежде всего из практических проблем, требующих огромных объемов вычислений. К настоящему времени машины позволили решить многие народно-хозяйственные задачи, которые без них были бы в принципе нерешаемы. Следует отметить двойственный характер причинно-следственной связи развития мощности компьютеров и сложности возникаемых задач. Взаимосвязь роста мощностей компьютеров и вычислительной сложности новых актуальных проблем можно сопоставить с проблемой «щит и меч», когда появление новых «наступательных средств» сразу порождает соответствующие «оборонительные» и наоборот. В этом отношении существует гипотеза: *«мощностей вычислительных средств будет всегда не хватать»*.

В 1960 г. президент Национальной федерации автоматизации и кибернетики Франции А. Дюкрок³ предсказывал проблемы, которые могут возникнуть в связи с повсеместным распространением информационных технологий⁴. Он писал: «Мы лишь весьма приблизительно угадываем перспективы кибернетической революции, ибо в прошлом нельзя найти никаких критериев для сравнения. <...> Мы находимся на пороге перемен в нашем обществе настолько же фундаментальных, как те изменения, которые были вызваны появлением книгопечатания, конвейерного производства или автомобиля. Эти изменения наступят, когда человек получит власть над информационным процессом» [3. С. 106].

Дюкрок оказался прорицателем, и первый информационный кризис разразился в начале 1970-х гг., проявившись на фоне экономического кризиса 1970-х в снижении эффективности информационного обмена в обществе:

- резко возрос объем публикуемых данных;
- общение между группами разных специалистов стало затруднено;
- возрос объем неопубликованной информации;
- выросла проблема межъязыкового обмена в мире.

Парадокс социальной коммуникации в условиях информационного кризиса: лавинообразный рост объемов информации, сопровождающийся информационным «голодом» (физиологическими ограничениями человека в восприятии и переработке информации и трудностями в выделении нужной информации из общего потока).

Самой актуальной и острой в мире проблемой стала проблема создания, сохранения и эффективного использования «информационных ресурсов». Произошло формирование еще одного самостоятельного вида общественного ресурса – информационного, позволяющего экономить большинство других ресурсов общества. Дальнейший прогресс общества в значительной степени связан сегодня с совершенствованием информационной инфраструктуры, эффективностью формирования, размещения и использования информационных ресурсов и продуктов.

Вторую информационную революцию связывают с появлением персонального компьютера (ПК). Создание ПК на время приглушило развивающийся информационный кризис, но не на долго.

Если первая революция охватывала средства передачи информации, то создание ПК решительно преобразовало способ формирования, организации и распространения знания. Это изобретение позволило массе людей приобщиться к профессиональным знаниям, общественному информационному богатству без помощи посредников и тем самым превратить профессиональные знания в лично-значимые. Интеллект человека, вооруженный ПК, стал инструментом, ориентированным на самопознание и саморазвитие. Персональная компьютерная вездесущность широко распахнула ворота в информационный век. Теперь осталось лишь обеспечить легкий и свободный доступ к информации по всему миру. Главное – уже никто не сопротивлялся его применению.

Революционное значение сетевых коммуникационных технологий уже осознано: «глобальная информационная инфраструктура», «информационная супермагистраль», «информация на кончиках пальцев» – это различные наименования третьей информационной рево-

³ А. Дюкрок – Albert Ducrocq.

⁴ В его время еще не было понятия «информатика», поэтому он говорил о кибернетике.

люции. Фактически эта революция интегрирует эффекты всех предшествующих, ибо создает технологическую основу объединения интеллектуальных способностей и духовных сил всего человечества. Как никакое предшествующее открытие или изобретение в мире информации и коммуникаций, сетевая информационная революция оказывает мощное воздействие на все сферы жизни общества: политику, экономику, культуру и, разумеется, на жизнь и безопасность людей.

Вот здесь как раз общество и натолкнулось на неразрешимый пока «второй» кризис информационных технологий.

Кризис информационных технологий. Привязывая традиционные циклы производства и потребления к информационной среде и перенося их в «Сеть» почти без изменений, современное общество топится в информации. Мутные потоки информации превысили индивидуальные и социальные возможности их фильтрации для полезного использования. Подавляющая часть циркулирующей в «Сети» информации человеку не нужна: требуются лишь результаты ее обработки. Низкосортное перепроизводство «информационных ресурсов» превратило «Сеть» в аморфную среду коммутатора, нагнетающего чрезмерные потоки информации.

По мнению Питера Друкера⁵, известного специалиста в области управления, в истории человечества прошли три информационных революции, а сейчас происходит четвертая. Друкер утверждает: «Неудовлетворенность высшего руководства данными, поставляемыми информационными технологиями, и привела в действие новую, следующую информационную революцию» [4].

«Сеть» необходима для общества как инструмент развития и система хранения «информации». Однако естественные механизмы самозащиты информационной среды не справляются с существующими потоками. Противостоять этой децентрализованной угрозе можно только с помощью создания единой распределенной интегрированной компьютерной среды.

С другой стороны, в «Сети» ширится новый феномен – глобальные компьютерные задачи и распределенные вычисления. Требования к интеграции компьютерных решений превысили возможности индустриального программирования.

Продвижение в «Сеть» массовых задач экономики и бизнеса, попавших в сферы интересов инвесторов, удалось в меньшей степени, чем ожидалось. Колоссальные инвестиции вместо дивидендов принесли глубокий спад в сфере информационных технологий. Быстрый откат индекса NASDAQ к уровню середины 1990-х гг. говорит об изначальной недооценке сложности глобальной интеграции информационных решений, хотя сами по себе информационные технологии продолжают интенсивно развиваться.

Попытки сходу взять глобальные задачи технологии индустриального программирования потерпели фиаско. Причины возникшего кризиса лежат не в экономике или политике, не в отсутствии инвесторов, не только в нехватке высококлассных специалистов, технических средств, передовых технологий, а в фундаментальных подходах к созданию информационных систем и программных продуктов, заложенных в используемую парадигму программирования. Во-первых, парадигмы, когда выражают законы природы, не подчиняются настроениям рынка и массового производства. Во-вторых, парадигма не должна давать повода назвать себя догмой.

Изначальная классическая модель фон Неймана уже более полувека служит теоретической основой архитектуры и программ универсального компьютера. Компьютеростроение материализовало ее в многомиллионных тиражах процессоров, программ, закрепило в триллионных оборотах информационного рынка. Но «сетевое» программирование потребовало создания качественно иного уровня интеграции глобального информационного пространства. На смену «объектам», захватившим умы программистов, должны прийти «процессы».

В том, что в рамках широко используемой парадигмы глобальная интеграция не состоялась, нет вины классической аксиоматики. Ее изначальные полномочия распространяются только на изолированный⁶ компьютер. Однако об ограничениях сфер действия «классики» благополучно «забыли». Индикатором неадекватности классических постулатов в «Сети»

⁵ Питер Друкер – Peter Ferdinand Drucker.

⁶ Может быть, даже очень «большой».

стал *информационный шум* «многоязычия», являющийся главным препятствием глобальной интеграции.

Деструктивная роль информационного шума «многоязычия» осознана еще в сюжете Вавилонского столпотворения. В отличие от социальной среды, где языки наряду с информационной несут и социально-психологическую нагрузку, компьютерную среду можно избавиться от «многоязычия». Полисемия (многозначность представления) – враг информационного обмена, фактор внесения искажения и ошибок (семантического шума) на пути передачи информации. Поэтому устранение многозначности является одним из наиболее важных направлений в разработке формальных подходов представления информации и знаний.

Общество уже преодолело первый кризис «многоязычия» информационных технологий в технической сфере: в конце 1980-х гг. был принят целый ряд стандартов ⁷.

Но наиболее серьезной проблемой является кризис в сфере представления информации. Важной деструктивной особенностью «Сети» является отсутствие единых общепринятых определений в сфере информационных технологий, когда речь идет об обработке «информации», прежде всего потому, что со времен Клода Шеннона на инженерном уровне произошло смешение понятий, объединение представлений об информации и данных или сигналах, кодирующих эту информацию. С легкой руки первопроходцев «информацией» стали называть, по существу, наборы данных.

До последнего времени, пока системы были относительно просты, отсутствие четкого разделения на данные и «информацию» не имело практического значения. Но с появлением сложных информационных систем, где функции распределены между человеком и машиной, где человек является активной составляющей, а также с развитием таких дисциплин, как управление знаниями, требуются более точные определения базисных понятий: «данные», «информация» и «знание». Преодоление этого кризиса возможно только с помощью новой, полной и замкнутой модели единого и математически однородного поля «компьютерной информации», не запрещающей программистам и пользователям строить удобные для себя языки и средства.

Интеграция в «Сети» достижима пока только для каждого отдельного случая ⁸. Индустриальное программирование никак не найдет общего языка с интеграцией. То, что надо одному, неудобно для другого. Имеющийся инструментарий вынуждает дробить глобальные задачи на части, решать их в отдельности (локализовать), а затем по мере необходимости интегрировать воедино.

Информационное общество. Основной лозунг своей предвыборной компании Жискард д'Эстен сформулировал так: «Информация является важнейшим стратегическим ресурсом и наибольший экономический и социальный успех сегодня сопутствует тем странам, которые активно используют современные средства компьютерных коммуникаций и сетей, информационных технологий и систем управления информационными ресурсами», – этим определив, что информационные технологии являются критическими технологиями развития общества [6]. Количество информации в современном обществе стремительно возрастает с каждым годом. По всей видимости, в ближайшем будущем информационные ресурсы станут основным национальным богатством.

Фундаментальная революция происходит на рубеже веков не в ядерной физике, микроэлектронике или химии, а в области «информатики». В экономически развитых странах осознан факт, что информационные технологии перестали выполнять чисто вспомогательные функции и стали важнейшим элементом деятельности государственных учреждений и частных компаний. Растущая зависимость промышленно развитых стран от источников информации (технической, экономической, политической, военной и т. д.), а также от уровня развития и эффективности использования средств передачи и переработки информации и привели к формированию принципиально нового понятия – «национальные информацион-

⁷ Стандарты взаимодействия открытых систем OSI, которые сейчас восприняты всеми производителями компьютерного и телекоммуникационного оборудования.

⁸ Наиболее ярким примером глобальной интеграции, но пока единственным, является единый распределенный каталог библиотек и международный стандарт библиографических описаний, созданный на основе протокола Z39.50 под эгидой Библиотеки конгресса США [5].

ные ресурсы». Возникновение в 1980-х гг. этого понятия было обусловлено растущей зависимостью промышленно развитых стран от источников информации (технической, экономической, политической, военной), а также от уровня развития и эффективности использования средств передачи и переработки информации.

Обобщая сказанное, можно выделить характерные черты и опасные тенденции информационного общества. Характерные черты:

- решается, хотя бы частично, проблема первого информационного кризиса – противоречие между информационной массой и информационным голодом;
- обеспечивается приоритет информационных ресурсов по сравнению с другими ресурсами;
- становление информационной экономики как главной формы развития общества;
- заложены основные принципы автоматизированной генерации, хранения, обработки и использования знаний;
- информационные технологии приобрели глобальный характер, охватив все сферы социальной деятельности человека;
- формируется глобальное единство человеческой цивилизации.

Опасные тенденции информационного общества:

- возрастающее влияние на общество средств массовой информации;
- все большее нарушение (или даже разрушение) посредством информационных технологий частной жизни людей или организаций;
- усложняющаяся проблема отбора качественной и достоверной информации;
- увеличение разрыва между разработчиками и потребителями информационных технологий до стратегически опасной величины;
- усиление проблемы адаптации части людей к среде информационного общества.

Трудно отыскать четкое указание на дату рождения кибернетики и информатики. Их черты как науки вырисовывались из многих смежных дисциплин, а становление связано с возникновением промышленности обработки данных. Небывалые темпы развития научных и технических достижений выдвинули в области обработки данных *информатику* на важнейшие позиции в современном обществе. Считается, что новые информационные технологии обеспечат занятость населения, эффективные результаты экономической деятельности и высокий уровень жизни. Однако в литературе нет четкого формального определения понятия «информация», поэтому если подходить формально, то, что такое *информационные технологии*, не совсем понятно.

Еще хуже дело обстоит в России. Учитывая прошлое негативное отношение к понятиям «кибернетика» и «компьютерная наука», вызванное политическими соображениями, термин «информатика» бесцеремонно отняли у скромной науки, называвшейся информатикой, но при этом ведавшей именно информацией, в основном научно-технической (во всем мире она называется *library science*)⁹.

Что такое информатика?

Информатика – это исключительно российское понятие, возникшее путем слияния понятий кибернетики и компьютерной науки, для определения сферы человеческой деятельности, связанной с повсеместным использованием вычислительной техники и телекоммуникаций.

Термином «информатика» мы привыкли называть области человеческой деятельности, связанные с разработкой, созданием, использованием, материально-техническим обслуживанием систем обработки информации, включая компьютеры, сети и программное обеспечение, а также организационные, коммерческие, административные и социальные аспекты компьютеризации (информатизации) – массового внедрения компьютерной техники во все области жизни людей.

⁹ Термин «информатика» (франц. *informatique*) родился в 1960 г., условно происходит от французских слов «*information*» (информация) и «*automatique*» (автоматизация) и дословно означает «информационная автоматизация», в научный обиход вошел в 1978 г.

Информатика изучает наши модельные представления об окружающей действительности – так называемые информационные модели, в которых на первое место выходит не портретное описание того или иного

явления (как в других науках), а описание информационных отношений, которые порождает это явление. Так же как и кибернетика, информатика является синтетической наукой, объединяющей целый ряд самостоятельных наук:

- вычислительные системы, программное обеспечение, методы машинной графики, анимации, средства мультимедиа;
- теория информации, изучающая процессы, связанные с передачей, приемом, преобразованием и хранением информации;
- математическое моделирование, методы вычислительной и прикладной математики и их применение к фундаментальным и прикладным исследованиям в различных областях знаний;
- методы искусственного интеллекта, моделирующие методы логического и аналитического мышления в интеллектуальной деятельности человека (логический вывод, обучение, понимание речи, визуальное восприятие, игры и др.);
- системный анализ, изучающий методологические средства, используемые для подготовки и обоснования решений по сложным проблемам различного характера;
- биоинформатика, изучающая информационные процессы в биологических системах;
- социальная информатика, изучающая процессы информатизации общества;
- телекоммуникационные системы и сети, в том числе глобальные компьютерные сети, объединяющие все человечество в единое информационное сообщество;
- разнообразные приложения, охватывающие производство, науку, образование, медицину, торговлю, сельское хозяйство и все другие виды хозяйственной и общественной деятельности.

Таким образом, информатика – это комплексная наука, занимающаяся разработкой методов и средств преобразования «информации» и организацией технологического процесса семантической (смысловой) обработки данных. Это определение позволяет, с одной стороны, подняться над статистической теорией информации К. Шеннона, где определяющая роль принадлежит знаковому представлению, а не смыслу сообщения, а с другой – позволяет не вступить на тернистый, но при этом непродуктивный путь философского анализа понятия «информация». Семантический же подход к информации прагматичен, он позволяет провести разделение между данными и информацией. Данные – это представление фактов и понятий в форме, пригодной для их передачи и интерпретации, а информация – это смысл, который ЧЕЛОВЕК приписывает данным на основании известных ему правил их представления. Объектами исследования информатики служат методы и средства, используемые для сбора, переработки, хранения, систематизации, поиска и распространения семантической информации.

Незрелость «информационных технологий». Взаимосвязь между «вычислениями»¹⁰ и составляющими информационную индустрию технологиями очень проста: она та же, что и в любой другой индустрии, в авиации, машиностроении, судостроении, – где угодно. Повсюду технологии обеспечивают строительный материал, строительные блоки для создания систем, обладающих конечными потребительскими свойствами – самолетов, станков, кораблей и т. д. Уровень зрелости индустрии определяется тем, насколько глубоко технологии скрыты от потребителя. Пользователи не обязаны понимать устройство бытовых электроприборов или автомобилей, они должны уметь ими ПОЛЬЗОВАТЬСЯ.

Сформулированный тезис показывает нам всю незрелость информационных технологий, несмотря на видимые успехи. Пока только некоторыми признаками зрелости обладают персональные компьютеры, но на корпоративном уровне применение этих технологий находится все еще на стадии становления.

Современные информационные технологии предоставляют исследователю мощный аппарат для «манипулирования данными», а не информацией. Данные, переведенные в электрон-

¹⁰ Под вычислениями здесь понимается процесс обработки информации с помощью компьютерных и сетевых устройств (компьютинг).

ную форму, приобретают новое качество, обеспечивая им более широкое распространение и эффективное использование. Однако применение информационных технологий должно основываться на использовании различных моделей (феноменологических, информационных, математических и др.). Как неоднократно отмечал А. А. Ляпунов, «нет модели – нет информации». Для возможности продуктивной работы нужны данные, превращенные в «информацию», представленную в виде «знаний» – «адекватного отражения действительности в сознании человека в виде представлений, понятий, суждений теорий» [7].

Для эффективного управления необходимо располагать «полноценной информацией»: информация должна быть не только полной, достаточной, но и своевременной, точной и непротиворечивой, а «информация», поступающая через различные по своей физической природе каналы, должна быть согласованной.

Информационные системы. Очевидно, автоматическая переработка «информации» возможна лишь при наличии ее описания с помощью некоторого алгоритма, т. е. при наличии формальной модели данных. Комплексы аппаратно-программных средств, позволяющие работать с данными, структурированными при помощи той или иной формальной модели, обычно называют информационными системами. Однако и здесь до сих пор присутствует путаница с понятиями «информационная система» (ИС) и «архитектура ИС». Она вовсе не безобидна и часто мешает на практике четко определить, что же является предметом разработки в конкретном проекте: ИС, только ее КСА (комплекс средств автоматизации) или система (АС) целиком. Для примера приведем два определения из государственных стандартов: «Информационная система (ИС): Комплекс, состоящий из процессов, технических и программных средств, устройств и персонала, обладающий возможностью удовлетворять установленным потребностям или целям»¹¹ и «Информационная система (ИС): система, предназначенная для сбора, передачи, обработки, хранения и выдачи информации потребителям и состоящая из следующих основных компонентов: программное обеспечение, информационное обеспечение, технические средства, обслуживающий персонал – Information system: The collection of people, procedures, and equipment designed, built, operated, and maintained to collect, record, process, store, retrieve, and display information»¹².

Основная проблема, связанная с построением информационных систем – незрелость теории *об информации*. Невозможно говорить о единой теории, связывающей воедино «данные», «информацию» и «знание», есть только множество подходов к понятию «информация» с философских или кибернетических позиций. Можно обнаружить десятки и сотни порой противоречащих друг другу определений того, что является информацией или знанием.

Существует правда «техническая» теория информации, по Шеннону, которую следовало бы называть «теорией передачи и хранения данных».

Существующую проблему фильтрации информации уже давно пытаются решить путем создания универсальных или специализированных информационно-поисковых систем. В результате опережающего развития технологий поиска по сравнению с методиками работы с семантической информацией образовался заметный разрыв между техникой работы с данными (поиском) и способностью работать с содержанием, заложенным в этих данных. Опираясь на интуицию, эксперты приходят к выводу о порочности нынешней ситуации, но о каком-либо серьезном переосмыслении проблем извлечения из данных информации пока речь не идет.

Стоит напомнить, что в 1950-е и 1960-е гг. в СССР науки *об информации* занимали весьма достойное место, хотя их развитие и было затруднено спецификой социального устройства общества. Произошедший в последующие годы разворот к работе с данными и узурпация названия «информатика» привели к тому, что многое из накопленного оказалось если не потерянным, то невостребованным, а культура работы с информацией была утеряна. Сегодня

¹¹ ГОСТ Р ИСО / МЭК 12207-99. Информационная технология. Процессы жизненного цикла программных средств // Госстандарт России. М., 1999.

¹² ГОСТ 34.003-90. Информационная технология. Комплекс стандартов и руководящих документов на автоматизированные системы. Термины и определения.

для большинства пользователей важнее потребление информационных сервисов, а не обеспечивающие его технологии.

Пути преодоления кризиса

Дальнейшее развитие работ в области информационно-поисковых систем и интеллектуализации информационных систем и систем поиска, несомненно, даст свои плоды, но очень дорогой ценой. Необходимо переосмысление самого подхода к представлению информации в «Сети».

Интеграция разнородных информационных ресурсов в единую среду и организация доступа к ним является одним из важнейших направлений развития современных информационных технологий. При этом наблюдается переход к исключительно распределенной схеме создания, поддержания, хранения ресурсов¹³ и в то же время стремление к виртуальному единству посредством предоставления свободного доступа к любым ресурсам сети через ограниченное число «точек доступа».

Парадигмы программирования. Одна из задач, направленных на преодоление кризиса, заключается в смене парадигмы программирования и устранении недостатков классической аксиоматики. В модели фон Неймана произвольные структуры данных формируются последовательностями (потоками) адресов к памяти (для чтения / записи элементов данных). Значения адресов в потоке задаются алгоритмическими предписаниями. Модель позволяет произвольным образом кодировать в потоках произвольные структуры данных. Следовательно, при создании алгоритмов имеются две степени свободы: структура данных и способ ее описания; обе они открыты программистам. По своему усмотрению они строят структуры данных и для каждой из них задают свои правила доступа, которые в дальнейшем плохо поддаются интеграции. Здесь и лежит основная причина воспроизводства информационного шума – избыточного и труднопреодолимого при интеграции множества в представлениях структурированной и слабоструктурированной информации. Решение проблемы – запрет на производство плохо совместимых структур данных путем создания однородного поля компьютерной информации посредством математической регламентации используемых структур.

Такая регламентация уже предлагалась ранее на основе методов декларативного программирования и исчисления древовидных структур. Деревья (или сети) – это рекурсивно-однородный структурный элемент компьютерной информации, представляющий программы и данные. Исчисление деревьев – это математически полные и замкнутые в множестве деревьев правила универсальной композиции (или декомпозиции). При программировании деревья свободно интерпретируются в терминах предметной области решаемых задач. С математической точки зрения, это исчисление «слабое», а потому не имеющее обременительных ограничений в отношении классов решаемых в нем задач.

На основе этой идеи в конце 1980-х гг. Международным телекоммуникационным союзом (ITU) была предпринята попытка построения единой универсальной информационно-справочной системы (RFC1487, X.500 Lightweight Directory Access Protocol. <http://rfc.net/rfc1487.html>).

Проблемы поиска. Идея структуризации информации уже формулировалась создателями WWW сервиса, которые считали, что создаваемая система должна не только поддерживать произвольные гипертекстовые ссылки, но предоставлять сквозной поиск по распределенным в сети базам данных и – что самое главное – быть многопользовательской, платформенно-независимой системой. Дальнейшее развитие технологии WWW немного подкорректировало начальные установки, и из системы доступа к разнородным ресурсам она превратилась во всемирную «презентационную» систему.

¹³ Эффективная эксплуатация информационных ресурсов возможна только в том случае, когда они постоянно поддерживаются авторами, т. е. на основе технологий использования распределенных информационно-вычислительных ресурсов, которые получили название GRID-технологий.

Пожалуй, главной проблемой «Сети» считается отсутствие какой-либо упорядоченности в петабайтах данных, раскиданных по его просторам. В полностью децентрализованной системе изначально не предусматривалось встроенных механизмов по каталогизации документов и обеспечению поисковых сервисов. И, поскольку само по себе существование информации не имеет смысла, если ее нельзя найти, различными компаниями создавались и создаются глобальные поисковые системы, осуществляющие индексацию всех найденных в «Сети» документов. При получении очередного документа такая система осуществляет его разбор, выделяет и сохраняет ключевые слова, по которым в дальнейшем этот документ может быть найден любым посетителем поискового сервера. Как правило, в роли ключевых слов выступают все слова, найденные в документе, за исключением предлогов, местоимений и прочих малоинформативных слов. Поисковые алгоритмы каждый год совершенствуются, улучшая их качество (повышая релевантность результатов поиска запросу): при поиске учитываются морфологические формы и / или синонимы ключевых слов, ограничивается «расстояние» между ключевыми словами, для сортировки результатов поиска используется понятие «веса» найденных в документе слов...

В целом, развитие таких систем, безусловно, оправдано. Однако, как уже отмечалось, это не дает решения проблемы, поскольку пока доминирующей идеей является улучшение средств отображения документов на стороне клиентов, а не их структуризация.

Службы каталогов. Спецификации X.500 по организации единой службы каталогов в «Сети», принадлежащие Международному телекоммуникационному союзу, были опубликованы в 1988 г. Служба каталогов представляет собой распределенную тиражируемую базу данных, где хранится логическое описание сетевых ресурсов. Конечная цель проекта X.500 была весьма амбициозна: создать распределенную международную систему каталогов, содержащую в себе все данные обо всех телефонных абонентах по всему миру и других сетевых ресурсах. После опубликования стандарта стало ясно, что задача построения глобального каталога практически неподъемна: технология оказалась слишком сложна, а конкретные ее реализации слишком ненадежны и малосовместимы между собой. Важным результатом этой работы стала очевидная актуальность новой задачи: создания средств хранения данных о реальных объектах в относительно простой, надежной и распределенной системе. Сетевые ресурсы организуются в иерархическую структуру, называемую деревом (в сети может быть несколько деревьев).

В 1993 г. группа из Мичиганского университета под руководством Тима Хауса опубликовала RFC 1487 с описанием LDAP¹⁴ – протокола, работающего непосредственно с TCP/IP и достаточно «легковесного», чтобы хорошо работать на обычных персональных компьютерах. Вначале предполагалось, что LDAP будет использоваться именно как дополнение к существующим продуктам X.500 для относительно маломощных компьютеров; в такой схеме клиент DAP является одновременно сервером LDAP. Производители каталогов X.500 поддержали этот подход и включили LDAP в свои продукты, что привело к резкому росту популярности каталогов масштаба организации. Каталоги LDAP, сохраняя логические стандарты X.500 (организация объектов в логическое дерево, хранение данных в виде «атрибут-значение» и т. д.), стали независимы от X.500 на уровне протоколов, полностью перейдя на стандарты Интернета. Еще одним существенным преимуществом технологии LDAP является четкое разделение между системой хранения и системой обмена данными.

Каталоги LDAP представляют собой универсальную систему хранения «метаданных» в «Сети» и целый набор профилей (правил) и схем данных для представления «информации». Схемой данных службы каталогов называется набор возможных и необходимых типов объектов и связанных с ними атрибутов с заданными способами взаимодействия между ними. Большое достижение служб каталогов по сравнению со службами доменов состоит в том, что их схемы являются расширяемыми. Иначе говоря, они позволяют регистрировать и вводить новые типы объектов или задавать новые атрибуты для уже имеющихся типов объектов.

Служба каталогов позволяет организовать гибкое управление правами доступа к сетевым ресурсам, к объектам, а также к отдельным атрибутам объектов. При рассмотрении службы каталогов важно знать, какие объекты выступают в качестве принципов безопасности.

¹⁴ LDAP – Lightweight Directory Access Protocol (<http://rfc.net/rfc1487.html>).

Принципалами безопасности называют объекты, которым могут быть даны права доступа к другим объектам, их свойствам или к сетевым ресурсам.

Службы каталогов имеют свойство, называемое наследованием прав доступа к ресурсам. Наследование прав доступа подразумевает следующее: если принципалу безопасности предоставлены определенные права на контейнер дерева, то он автоматически получает такие же права на все входящие в контейнер объекты, среди которых могут быть, в свою очередь, вложенные контейнеры и т. д. Ограничение прав доступа на вложенные объекты осуществляется с помощью фильтров наследования. Применение свойства наследования прав доступа совместно с фильтрами наследования позволяет очень гибко распределять обязанности по администрированию сетевой среды.

Распределенные вычисления. Второе направление связано с реализацией схем распределенных вычислений и доступа к распределенным ресурсам в «Сети». Здесь можно выделить два направления: одно – «снизу вверх» (от практики), другое – «сверху вниз» (от теории). Пример первого направления – Grid [8], подход к интеграции машинных ресурсов «Сети» для распределенных вычислений и систем. Многочисленные и разнородные версии Grid основаны на классической парадигме программирования, потому органично вписываются в арсенал глобального программирования. Это путь постепенного наращивания средств индустриального программирования в рамках классической парадигмы. Интеграция осуществляется силовым преодолением информационного шума, путем формирования общих вычислительных сервисов.

Второе направление, фактически заложенное в архитектуру служб каталогов, связано с разработкой новой парадигмы математически однородного поля компьютерной информации. В этой парадигме проблемы программирования, интеграции и масштабирования компьютерных решений и распределенных процессов перестают зависеть от технических особенностей компьютеров и сетей. Регулярный инструмент глобального программирования строится сначала на теоретическом уровне в виде математически полной и замкнутой модели. Затем алгоритмы функционирования этой модели программно эмулируются в ресурсах «Сети».

В математически однородном поле компьютерной информации интегрируются не ресурсы машинной среды («объекты»), а решаемые в нем задачи («процессы»). Здесь еще предстоит построить математически замкнутую модель представления информации и процессов, необходимую для реализации универсального ядра глобально распределенных вычислений и систем.

Новый подход подразумевает и новую стратегию: проблемы переустройства «Сети» выходят за рамки какой-либо корпорации и должны опираться уже не на законы рыночной конкуренции, а на равно обязательные для всех фундаментальные свойства и законы развития компьютерной информации.

В заключение уместно отметить постоянно развивающуюся тенденцию к интеграции уже существующих информационных ресурсов. Одним из необходимых условий для этого является каталогизация информационных ресурсов и документов на основе существующих открытых стандартов.

Список литературы

1. Шокин Ю. И., Федотов А. М., Гуськов А. Е., Жижимов О. Л., Столяров С. В. Электронные библиотеки – путь интеграции информационных ресурсов Сибирского отделения РАН // Вестн. Казах. нац. ун-та. 2005. № 2. Спец. вып. С. 115–127.
2. Колмогоров А. Н. Теория информации и теория алгоритмов. М.: Наука, 1987.
3. Дюкрок А. Физика кибернетики // Кибернетика ожидаемая и кибернетика неожиданная. М.: Наука, 1968.
4. Хаас Э. Э. Лучшие идеи Питера Друкера. СПб.: Питер, 2008.
5. Жижимов О. Л., Мазов Н. А., Федотов А. М., Шокин Ю. И. Сервер ZooPARK как сервер для построения распределенных информационных систем // Информационные технологии в высшем образовании. 2005. Т. 2, № 1. С. 53–67.
6. Д'Эстен В. Ж. Французы. Размышления о судьбе народа. М.: Ладомир, 2004. 248 с.

7. Ляпунов А. А. О соотношении понятий материя, энергия и информация // Ляпунов А. А. Проблемы теоретической и прикладной кибернетики. Новосибирск: Наука, 1980. С. 320–323.

8. *The Grid: Blueprint for a New Computing Infrastructure* / Eds. I. Foster, C. Kesselman. San Francisco, CA: Morgan Kaufmann Pub., 1999.

Материал поступил в редколлегию 13.08.2008

A. M. Fedotov

Problems of Information Technologies

The paper is devoted to problems and ontology's history of information technologies.

Keywords: information technologies, information resource, history, network, information revolution, programming paradigms, search.