

Безбумажная информатика

Ф. Каплун,
Г. Аванесян

Зародившись как самостоятельная область исследований в начале XX века, информатика

менее чем за столетие преодолела дистанцию огромного размера. Из теоретической дисциплины, занимающейся только научной информацией, она превращается в важнейшую область науки и техники, внедряющуюся во все сферы деятельности человека. Новые горизонты перед ней открывает прогресс в электронике и вычислительной технике. Публикуемая ниже статья посвящена настоящему и будущему информатики.

Обращаясь к технической литературе начала 70-х годов, можно составить некоторое собирательное определение информатики как научной дисциплины, занимающейся исследованием общих свойств научной информации и закономерностей, связанных с процессами научных коммуникаций. А ведь совсем не так давно предмет ее ограничивался принципами и структурой реферирования научной документации без изучения содержания этих документов. Это на первый взгляд очень схоже с тем, что изучает библиотекведение. Действительно, на начальных этапах сферы интересов этих дисциплин были весьма близки.

Этапными моментами в становлении новой области знаний стали международные конференции по научной информатике, прошедшие в Лондоне (1948 год) и Вашингтоне (1958 год). К тому времени необходимость серьезных исследований с целью создания автоматизированных систем сбора, хранения и поиска научной информации ощущалась достаточно остро. С одной стороны, «информационный потоп», о котором сегодня так много говорят, уже становился реальностью и лавиную долю своего времени служащие начали тратить на работу с захлестывающими их потоками информации. С другой, прогресс в полупроводниковой электронике и вычислительной технике (в 1946 году в США создана первая цифровая вычислительная машина ENIAC) сделал технически реализуемым разделение труда между машиной и человеком в области переработки информа-

ции. ЭВМ брала на себя часть рутинной работы, оставляя человеку решение более творческих задач. По мере совершенствования информационных технологий водораздел между функциями ЭВМ и человека все более смещается в область творчества.

Автоматизированное, а зачастую и автоматическое выполнение ряда задач привело к тому, что огромные объемы информации перестали документироваться на бумаге. ЭВМ, обладая свойством практически неограниченно долго хранить в памяти данные, стала создавать свою, «электронную», бумагу. Настала эра безбумажных технологий — автоматизированных банков данных и интерактивных систем.

Обобщенная модель информационной системы может быть представлена в виде комплекса компонент, формализующих протекающие процессы (рис.1). Начальным этапом следует считать этап рецепции информации, на котором происходит выделение информации по одному или нескольким информативным признакам. Назовём их рецептивными и обозначим как X, а сам процесс рецепции поставим в функциональное соответствие датчикам восприятия.

Чтобы внести необходимую ясность, заметим, что в зависимости от принимаемой глубины формализации может быть различным и описание самих датчиков. Рассматривая датчик как отдельный узел, приходится давать системное описание. В этом не сложно убедиться, взяв для при-

мера такой биологический объект, как нейрон. Многочисленные дендриты нейрона преобразуют возмущения из внешнего мира в последовательность электрических импульсов постоянной амплитуды, закладывая информацию в их частоту. Налицо совершение серии операций, приводящих к рецепции. Но что же тогда считать информативным признаком? Если в качестве датчика принять нейрон, реагирующий на тепловые воздействия, то само собой разумеется, что информативным признаком является температура. С другой стороны, на выходе нейрона как рецептивной системы мы имеем последовательность частотно-модулированных импульсов. Значит, справедливо было бы ввести классификацию информативных признаков в зависимости от этапов информационного процесса и задач формализации. Будем далее считать информативные признаки, служащие непосредственно для рецепции, рецептивными, а используемые для выбора уже поступившей в систему информации — классифицирующими.

Вторая компонента информационной модели — система селекции по информативным признакам. Этап селекции в частном случае может отсутствовать, например при обработке сигналов с одним информативным признаком или характерной особенностью рецептивной системы воспринимать ограниченное число параметров. В общем же случае при создании современных автоматизированных систем сбора и интерпретации информации селективный отбор

вытекает уже из постановки задачи, поскольку известный аналитический подход предусматривает поступление на соответствующие входы информационной системы двух потоков: информации об объектах (классах объектов) и информации, определяющей информационную потребность абонентов. При селекции данных по информативным признакам в качестве меры удобно использовать условную энтропию H :

$$H(K/X) = -\sum p(x) \sum P(k/x) \lg P(k/x),$$

где K — множество классов объектов, исследуемых по информативным признакам X ;
 k — порядковый номер класса;
 x — сигнал в пространстве признаков X , появление которого задаётся плотностью вероятности $p(x)$;
 $P(k/x)$ — апостериорная вероятность класса k при условии наблюдения сигнала x .

Чем ниже значение условной энтропии H , тем более достоверным оказывается результат определения класса, а следовательно, более эффективно решается задача поиска информации. Случай, когда $H(K/X)=0$, отвечает ситуации максимально достоверного ответа на информационный запрос. Из приведённой зависимости следует также, что на минимизацию H влияет правильный выбор набора информативных признаков X с точки зрения их

пригодности для распознавания соответствующих классов.

Селекция данных может происходить на различных этапах, выполняя, соответственно, разные функции, но подчиняясь при этом единым правилам. Селекция первичного уровня, т.е. после процедуры рецепции, необходима для формирования специального банка данных по общему критерию, составляющего "память" информационной системы. Селекция вторичного уровня призвана выделить из уже сформированного банка необходимую информацию по запросам абонентов. Однако до поступления на хранение ее необходимо интерпретировать в форме, пригодной для дальнейшего обращения. Вопросам интерпретации информации посвящено немало работ в области радио- и вычислительной техники. Мы же остановимся на ключевых моментах, характерных для информатики.

Исходя из общей и обоснованной тенденции обработки сигналов в дискретизированном виде, к начальной стадии интерпретации отнесём преобразование аналоговой информации $x(t)$ в цифровую, выражая преобразование через оператор $\Phi_n(x(t))$. После преобразования с точностью, задаваемой числом уровней квантования m и частотой дискретизации f_d , имеем сигнал, представляемый параметрической функцией $\tilde{X}(t, f_d, m)$. Далее с целью

рационального использования памяти системы информацию сжимают, что устраняет ее избыточность, но, с другой стороны, снижает помехозащищенность. Поэтому вопрос должен решаться в каждой ситуации строго индивидуально, исходя из конкретной помеховой обстановки и способов коммуникации с абонентами.

На этапе интерпретации информация об объекте может быть заменена на информацию о модели объекта. С одной стороны, это характеризует степень интеллектуализированности системы, а другой, расширяет множество действий, совершаемых над преобразованными сигналами. Таким образом, процесс интерпретации можно представить композицией операторов Φ_n :

$$\Phi_n = \Phi_{сж}(\Phi_m(\Phi_p(\tilde{X}))),$$

где $\Phi_{сж}$, Φ_m и Φ_p — операторы сжатия, моделирования и преобразования, соответственно.

Интерпретированную информацию записывают в память системы, откуда ее извлекают для участия в конечном процессе — коммуникации. На этом этапе для поиска требуемой информации абонент обращается к системе селекции вторичного уровня, на долю которой приходится большая часть нагрузки. Система селекции второго уровня в практических приложениях может быть представлена различной комбинацией как технических устройств, так и взаимоотношений "человек-машина". Самая распространённая форма — клиент-сервер (файлов или приложений), а наиболее целесообразная форма хранения информации — распределённые базы данных. Последнее за счёт распределения ресурсов позволяют повысить доступность информации.

Информатизация общества, обусловленная, с одной стороны, совершенствованием средств вычислительной техники и, с другой, увеличением потоков различной информации, привели к чёткому делению создаваемых информационных систем на управляющие и справочно-поисковые. При анализе обеих систем за счёт формализации выполняемых

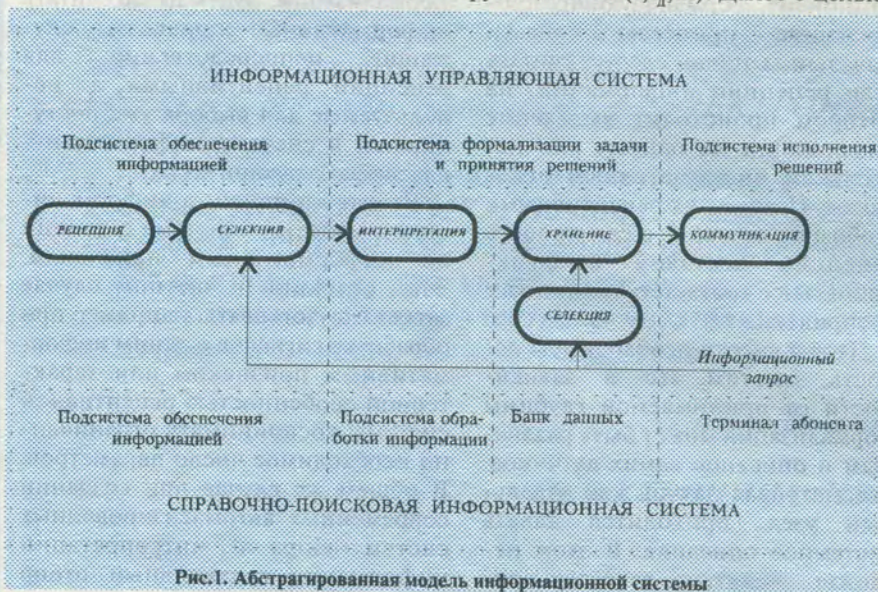
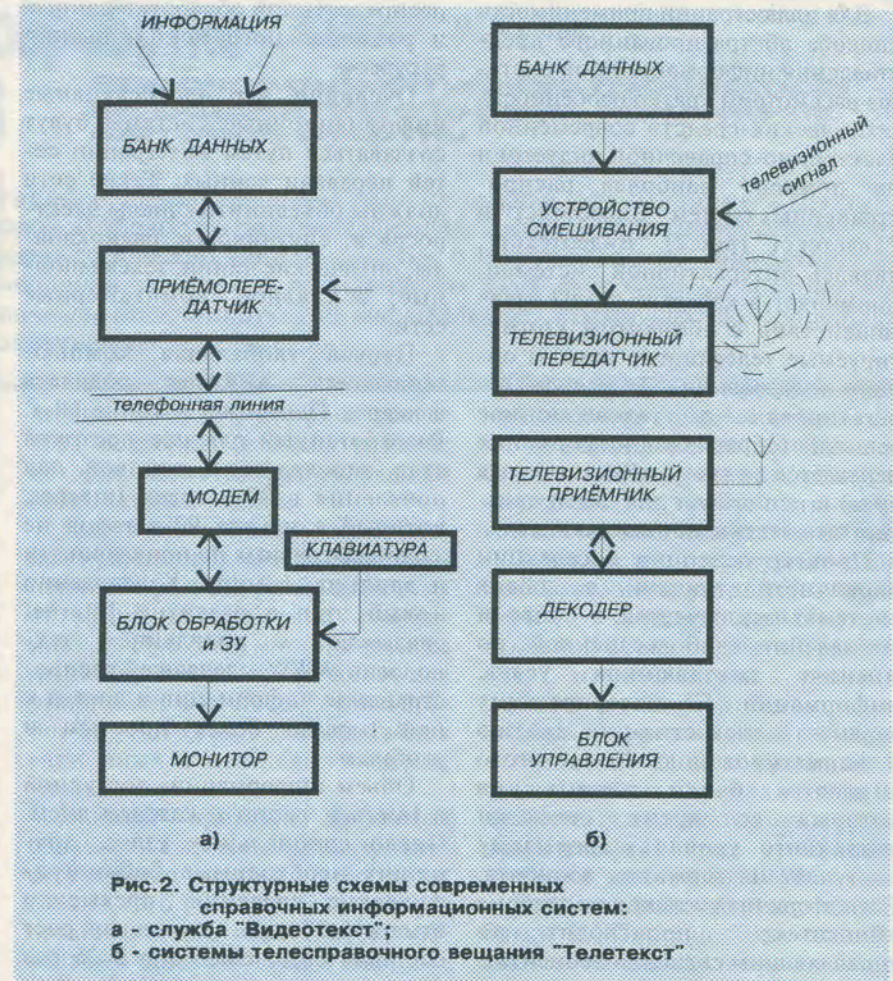


Рис.1. Абстрагированная модель информационной системы

действий может использоваться единая абстрагированная модель информационной системы, показанная на рис.1. Задача управления сводится к разработке структуры, включающей подсистемы обеспечения информацией, формализации задачи, принятия и, в качестве управляющего звена, исполнения решений.

Характерная особенность современных информационных управляющих систем — автоматизация решения аналитических проблем, требующих обработки огромных массивов информации, распределённой в пространстве. В связи с этим управляющая система превращается также в физическую распределённую систему, которая отличается от классических систем автоматического уровня тем, что в основу её построения заложены основные принципы функционирования систем информатики. Техническая структура справочных информационных систем также строится на основе рассмотренной модели (рис.1) и содержит подсистемы обеспечения информацией, ее обработки, банк данных и абонентский терминал. Отличие двух систем заключается в способе формирования информационных запросов и использования ответов. Если в первой системе информационные запросы формируются автоматически внутри самой структуры, то во второй информация запрашивается абонентом. Единство систем, состоящее в обязательном выполнении комплекса алгоритмических действий, а в техническом смысле — в аппаратной унификации, позволяет говорить о новой универсальной структуре, реализуемой сегодня на базе персональных ЭВМ, объединенных в единую сеть.

Обретя столь мощный инструмент, человечество получило возможность скачкообразно повысить эффективность и производительность процесса управления. Объем данных, собираемых для управления объектом независимо от его природы (социальный, биологический или технический), может быть эквивалентен тысячам бумажных страниц. После электронной обработки они участвуют в при-



нятии и исполнении решений, выдаче справок, формировании запросов и др. Традиционные механизмы управления, не имея резервов для повышения произво-

дительности, передают часть работы, совершаемой человеческим мозгом, вычислительной машине, оставляя человеку решение трудно формализуемых задач.



Рис.3. Прогноз продаж серверов (аппаратных и программных средств) сети Internet и Intranet (в млрд.долл.)

Для иллюстрации применимости способа абстрагированного представления информационной системы рассмотрим кратко особенности технических средств современной электронно-справочной индустрии на примере наиболее распространенных систем "Видеотекст" и "Телетекст" (рис.2). "Видеотекст", являясь интерактивной системой, применяется в качестве среды взаимодействия с абонентом коммутируемые телефонные каналы общего пользования. "Телетекст" агрегирован в телевизионное вещание (справочная информация передается во время обратного хода луча) и использует для связи стандартные телевизионные каналы.

Процесс рецепции и селекции первичного уровня в обеих системах заключается в сборе и обновлении наиболее ценной, по мнению поставщиков услуг, информации. Ее интерпретация сводится к дискретизации, сжатию и форматированию. После этого создаются банки данных, из которых во время селекции вторичного уровня информация поступает на терминал абонента. Если вторичная селекция в системе "Видеотекст" происходит по управляющим сигналам абонента в месте расположения банка данных, например на справочной станции, то в системе "Телетекст" селекция осуществляется в терминале (телевизоре) при помощи декодирующего устройства. Это хотя и примитивный, но все-таки пример

распределенной обработки данных и рационального использования ресурсов.

Очевидно, что перспективные информационные системы будут создаваться путем интеграции сетей передачи данных. Такие сети должны объединить каналы электросвязи, спутниковые радиоканалы, оптические линии, телевизионные, а главное, компьютерные сети.

Первая глобальная компьютерная сеть — APRAnet — родилась в недрах Пентагона в конце 60-х. Разработанный для нее протокол стал впоследствии основой для появления в 80-х годах Internet, который с начала 90-х годов не знает себе равных по популярности и влиянию в мире. Качественно новый этап в развитии Internet связывают с появлением технологии WWW, сделавшей распространение информации и доступ к ней гораздо более простым и удобным.

Объем информации, доступной в Internet, растет с каждым днем. Число специальных узлов, хранящих необходимые библиографические сведения, превысило отметку в 10 млн., а их годовой рост составил 85%. Ежегодно в 25 раз увеличивается и число Web серверов. По данным аналитической фирмы Zona Resaerch (США), за период 1995—1999 годов объем продаж Web серверов (аппаратных и программных средств) для сети Internet увеличится с 622 млн.

до 2,16 млрд. долл. Еще более высокими темпами будет расти объем продаж Web серверов Intranet — внутрифирменной компьютерной сети, развивающейся на базе Internet с 1996 года (рис. 3).

По прогнозам, уже в будущем году Internet станет столь же распространенным на Западе, как и телефон. Доступ к этой глобальной компьютерной сети по телефону сегодня стал привычным для ее пользователей. Завтра столь же привычным будет телефонный разговор по Internet. Специалисты считают целесообразным использовать для передачи данных каналы кабельного телевидения. Развитие Internet идет по всем направлениям. На рынке уже появляются программы просмотра, объединяющие WWW, электронную почту, удаленный вход и другие сетевые сервисы.

Постоянное совершенствование сети говорит о высокой жизнеспособности и потенциале Internet, которому предстоит реализоваться в следующем столетии. Однако многие ученые, видящие будущее человечества в формировании "информационного общества", объединенного коллективным разумом, серьезно говорят о том, что возможностей сетей типа Internet не хватает уже сегодня — необходим регламентированный отбор и классификация информации. Задача эта чрезвычайно сложна, но и ее решение не за горами. А дальше новые и новые шаги к горизонту.

Представляем авторов номера

КАПЛУН Фердинанд Вильевич. Доктор экономических наук, академик Международной академии информатизации, Заслуженный изобретатель СССР, автор 360 изобретений в области оперативно-тактического вооружения, исследования малого и среднего космоса и средств для носителей информации. В настоящее время генеральный директор издательства "Вся Москва — Информ".

Новый центр НИОКР в Йокогаме

Южнокорейская фирма Samsung Electronics открыла в Японии (г. Йокогама) новый центр НИОКР. Здесь будут проводиться научные исследования по 12 направлениям, включая компоненты и системы комплексной обработки данных, электронные приборы, средства связи, системы управления производством, программное обеспечение. В настоящее время штат большинства центров НИОКР южнокорейских фирм составляет 20—30 исследователей, ведущих работы в какой-либо одной области. В новом Йокогамском центре НИОКР будет занято несколько сот человек.

Кроме того, фирма Samsung ведет работы по расширению действующего в Японии с 1992 года Института передовых технологий, на что ассигновано 130 млн. долл. Окончание работ — май 1997 года. В 1998 году Samsung планирует также открыть центры НИОКР в Токио и Осаке.

Electronic Components, 1996, v. 17, N11, p.30